Universitá degli Studi di Bari
Master: "Metodi quantitativi e informatica
a supporto delle decisioni economiche"

Relazione del Progetto finale

Surface Reconstruction

Dr. Giuseppe Romanazzi

ANNO ACCADEMICO 2001/2002

Indice

1 Descrizione del problema **4**

1.1 Introduzione 4

1.2 Definizioni di base 5

1.2.1 Intensità di luminosità 5

1.2.2 Digitalizzazioni 5

2 Regolarizzazione **6**

2.1 Interpretazione stocastica del termine di regolarizzazione attraverso una stima MAP 8

2.2 Convessità e GNC 9

2.3 Mean Field Annealing 12

2.4 Uso dei filtri 13

3 Weak Continuity **16**

3.1 Introduzione dei line process 18

4 Line process e Regolarizzazione robusta **20**

4.1 Robust regularization 21

4.1.1 Applicazione delle norme robuste nell'Hamiltoniana 23

4.2 Line Process 25

4.3 Equivalenza 26

5 Uso dei parametri α e λ **29**
5.1 Limite di sensitività 29
5.1.1 Discontinuità interagenti 32
5.2 Discontinuità spurie generate dalla digitalizzazione . . . 35
5.3 Rumore e discontinuità spurie 37
5.4 Modello della weak membrane 38
5.5 Comportamento della ricostruzione 41
5.6 Modello adattivo della weak membrane 42

6 Algoritmo GNC di Blake & Zisserman **46**
6.1 Determinazione del funzionale convesso 48
6.1.1 Caso della weak membrane 49
6.1.2 Sufficienza del primo passo 51
6.2 Secondo passo dell'algoritmo GNC 53
6.3 Correttezza dell'algoritmo GNC 55
6.3.1 Discontinuità interagenti 59
6.3.2 Rumore 60
6.4 Individuazione dell'algoritmo 60
6.4.1 Scelta del parametro di rilassamento ω 63
6.4.2 Implementazione dell'algoritmo 63

7 Algoritmo GNC per il modello weak membrane **68**
7.1 Implementazione algoritmo GNC 68
7.2 Codice per l'algoritmo GNC 71
7.3 Descrizione del codice "GNC.c" 72
7.4 Applicazione del codice sull'immagine "lena" 76

A Appendice **80**
A.1 Discretizzazione dell'energia 80
A.2 Condizioni per la convessità 81
A.2.1 Un'altra espressione del modello weak membrane 84

A.3 Schema di passaggio da norme robuste a funzioni di penalità 86
A.4 Formula di Eulero Lagrange 87
A.5 Variabili aleatorie gaussiane 88

Bibliografia **90**

Codice "GNC.c" **91**

Capitolo 1

Descrizione del problema

1.1 Introduzione

In questa relazione tratteremo il problema della ricostruzione di immagini bidimensionali di input. Perché c'è la necessità di ricostruire le immagini?

La risposta è che le immagini, anche se acquisite dai più moderni sensori, possono essere contaminate da una varietà di sorgenti aleatorie di rumore che modificano l'aspetto dell'immagine reale. Dove per rumore si intende la variazione stocastica dell'immagine, che è cosa diversa dalle distorsioni deterministiche della stessa dovuta alle ombre o alla mancanza di messa a fuoco. L'importanza di rielaborare l'immagine di input per produrre un risultati senza rumore è fondamentale nel campo della neuroanatomia, neurofisiologia dove i recettori visivi biologici possono creare del rumore, ostacolando così l'analisi medica.

1.2 Definizioni di base

1.2.1 Intensità di luminosità

Un immagine può essere definita come una funzione in due variabili $u = u(x, y)$ dove u è l'intensità della luminosità dell'immagine nel punto di coordinate (x, y). La luminosità di una data immagine può essere espressa in forma di numeri interi, reali o essere quantizzata con un valore complesso. Nell'algoritmo associato a questa relazione si usano segnali reali anche se essi poi corrisponderanno, nella mappa delle luminosità, usata per rappresentare l'immagine, a 256 livelli di grigio interi.

1.2.2 Digitalizzazioni

I principali calcoli analitici sulle funzioni di intensità che affronteremo sfruttano la continuità del dominio Ω. Ma nel campo computazionale per applicare i risultati dedotti dalla teoria, il tutto si deve ridurre a considerare l'immagine campionata o meglio digitalizzata. Cioé dalla $u(x, y)$ definita in uno spazio 2D continuo Ω, si deduce un immagine digitale $u(m, n)$ con $m, n \in \mathbb{N}$ attraverso un processo di digitalizzazione (si veda [1]). L'immagine in questo processo viene divisa in M righe e N colonne. La zona corrispondente all'intersezione di una riga con una colonna è detta *pixel* e viene individuata dalle sue coordinate (m, n) con $1 \leq m \leq M$ e $1 \leq n \leq N$. Nel nostro programma vengono usate immagini di dimensioni quadrate del tipo $M \times N$ con $M = N = 512$. L'immagine digitalizzata sarà allora del tipo $u = u(m, n)$ con $u(m, n)$ che è uguale al valore intero più vicino alla media delle luminosità dei punti all'interno del pixel. Tale processo che rappresenta l'intensità di luminosità mediante dei valori interi riferiti ad una scala di differenti livelli di grigio è detta quantizzazione.

Capitolo 2

Regolarizzazione

I segnali corrotti da rumore ed in particolare le immagini possono essere ricostruite mediante l'Hamiltoniana che esprime l'energia $E(u, d)$ di un sistema (una membrana) in funzione delle forme che tale membrana assume e delle misure a disposizione. Questa energia si esprime come somma di due termini

$$E(u, d) = D + S = \int_\Omega ((u(s) - d(s))^2 \, ds + \int_\Omega f(u(s)) \, ds \qquad (2.1)$$

dove Ω è il dominio della ricostruzione $u : \Omega \to \mathbb{R}$ dell'immagine, $d : \Omega \to \mathbb{R}$ individua le misure a disposizione, $f = f(u)$ è la *funzione di stabilizzazione (o di regolarizzazione).* e ds è l'elemento infinitesimo dell'area Ω.

La membrana (o superfice) ricostruita finale $\widetilde{u}$ è ottenuta dalla minimizzazione dell'energia $\min_u E(u)$ al variare delle u in uno spazio di soluzioni convesso o meno.

Il primo termine della (2.1) D porta, nel procedimento di minimizzazione, a soluzioni che approssimano le misure iniziali $d(s)$ mentre il secondo S, detto *termine di regolarizzazione*, forza la regolarità delle soluzioni.

Thikonov[1], fu il primo ad introdurre come funzione di regolarizzazione la somma dei quadrati delle derivate

$$f(u) = u_x^2 + u_y^2.$$

Allora il termine di regolarizzazione

$$S = \int_\Omega u_x^2 + u_y^2 \; dxdy$$

riformula il problema di minimizzazione inserendo nel funzionale dell'energia la conoscenza a priori della superfice che si vuole costruire: che deve essere regolare perchè con gradienti limitati. Inoltre tale formula porta ad uno spazio di soluzioni convesso cioé in questo caso il problema della minimizzazione ammette un'unica soluzione.

Non sempre però la superfice da ricostruire è regolare, infatti le immagini hanno, in genere, parecchi punti di discontinuità corrispondenti a quelli in cui la luminosità varia con un grande salto ed allora Geman & Geman[2] introdussero un campo di discontinuità nella ricostruzione, cioé proposero il modello della weak membrane (membrana fragile) che venne formulato nel caso continuo (non discretizzato) da Mumford e Shah[3]. Tale modello discontinuo, che sfrutta la weak continuity (vedi successivo capitolo), venne incorporato direttamente, mediante un algoritmo numerico, nella risoluzione del problema di minimizzare $E(u, d)$ da Blake e Zisserman [2]; il loro studio d'ora in poi verrà indicato con B&Z e sarà esaminato più in dettaglio nei prossimi capitoli.

---

[1] A.N. Thikonov and V.Y. Arsenin *"Solution of ill posed problems"*, V.H. Winston & Sons, John Wiley, 1977.

[2] S. Geman and D. Geman *"Stochastic Relaxation, Gibbs Distribution, and the Bayesian Restoration of Images"*, IEEE PAMI Vol.6, No. 6, pp. 721-741, April 1984.

[3] D.Mumford and J.Shah *"Optimal Approximation by Piecewise Smooth Functions and Associated Variational Problems"* Comm. on Pure and Applied Math, Vol. 8, No. 4, 1990.

Il funzionale di regolarizzazione f può essere interpretato anche come misura di entropia o in termini di stima Bayesiana usando uno schema MAP (Maximum A Posteriori) [4]. Qui di seguito c'è un breve cenno a quest'ultima interpretazione.

2.1 Interpretazione stocastica del termine di regolarizzazione attraverso una stima MAP

Consideriamo un modello di rumore $P(d|u)$, che viene rappresentato in questo modo perché la probabilità di determinare il rumore n è uguale a quella di ottenere i segnali $d = u + n$ data la superfice reale u. Sia inoltre $P(u)$ la distribuzione, nota a priori, delle superfici da costruire allora la probabilità a posteriori di ricostruire u è data dalla *formula di Bayes*

$$P(u|d) = \frac{P(d|u)P(u)}{P(d)}$$

con $P(d)$ una costante nota e dove vale $P(a|b) \equiv \frac{P(a\&b)}{P(b)}$. Ora nella pratica la distribuzione nota delle u è esprimibile in termini delle loro derivate ∇u cioé è del tipo $P(\nabla u)$. Inoltre usando un modello di rumore additivo non correlato gaussiano per $P(\nabla u)$ si ha che la funzione di regolarizzazione, espressa in funzione di ∇u, è

$$g(\nabla u) = -\log(P(\nabla u))$$

infatti essa rende regolare le u se viene minimizzata. Siccome la funzione $-\log$ è monotona decrescente la ricerca del minimo dell'energia

$$\sum (u - d)^2 + g(\nabla_u)$$

[4] D. Kazakos and P. Papantoni-Kazakos *"Detection and Estimation"* Computer Science Press, W.H. Freemann and Company, New York, 1990.

corrisponde a massimizzare la probabilità a posteriori $P(\nabla u|\nabla d)$, infatti $g(\nabla u)$ è minima se $P(\nabla u)$ è massima ossia se $P(\nabla u|\nabla d)$ è massima. Quindi usando una stima MAP il problema dell'energia si riduce a massimizzare la probabilità a posteriori.

2.2 Convessità e GNC

In entrambi i casi di regolarizzazione (deterministica o probabilistica) le proprietà di ricostruzione e di convessità, che ci garantiscono la ricerca del minimo assoluto senza cadere in una concavità relativa ad un minimo solo relativo, dipendono dalla funzione di regolarizzazione. Thikonov usava una funzione di regolarità parabolica mentre B&Z una parabola troncata che non garantiva la convessità. Allora per trasformare lo spazio di soluzioni in uno convesso sono stati introdotti vari algoritmi GNC (Graduated Non Convexity). Nei quali attraverso l'introduzione di parametri lineari si trasforma lo spazio delle soluzioni in uno convesso. In questo nuovo spazio si determina la u che minimizza l'energia e facendo variare di poco e in maniera lineare i parametri introdotti si cerca di ritornare al problema iniziale non convesso applicando, passo dopo passo, un algoritmo di ricerca del minimo che parte dal minimo del passo precedente. Analizziamo ora l'algoritmo GNC applicato all'Hamiltoniana (2.1) con la differenza che la funzione di regolarizzazione è espressa in funzione del gradiente della ricostruzione

$$E(u,d) = D + S = \int_\Omega ((u(s) - d(s))^2 \, ds + \int_\Omega g(\nabla u(s)) \, ds. \qquad (2.2)$$

Il primo passo di un algoritmo GNC consta allora nell'individuazione di una funzione g^* che rende il funzionale dell'energia E^* convesso. Mentre nel secondo passo partendo da $E^{(1)} \equiv E^*$ si costruisce una successione di funzionali dell'energia $E^{(p)}$ che approssimano sempre meglio al diminuire di p il funzionale di partenza (2.2). E nei quali si deter-

minano le ricostruzioni $u^{(p)}$ che minimizzano l'energia applicando un metodo di discesa che parte da $u^{(p+1)}$. Per far si che la successione delle $u^{(p)}$ porti alla soluzione del problema iniziale si deve costruire opportunamente il funzionale convesso e gli altri funzionali $E^{(p)}$, infatti essi devono individuare uno sviluppo di deformazione del funzionale di partenza che concentri la convessità nella concavità del minimo assoluto (Figura 2.1); altrimenti si avrebbe nel secondo passo, in cui si ha il passaggio $E^{(1)} \rightarrow E^{(0)}$, l'individuazione di un minimo solo locale di $E \equiv E^{(0)}$. Una condizione che garantisce la convessità per un funzionale

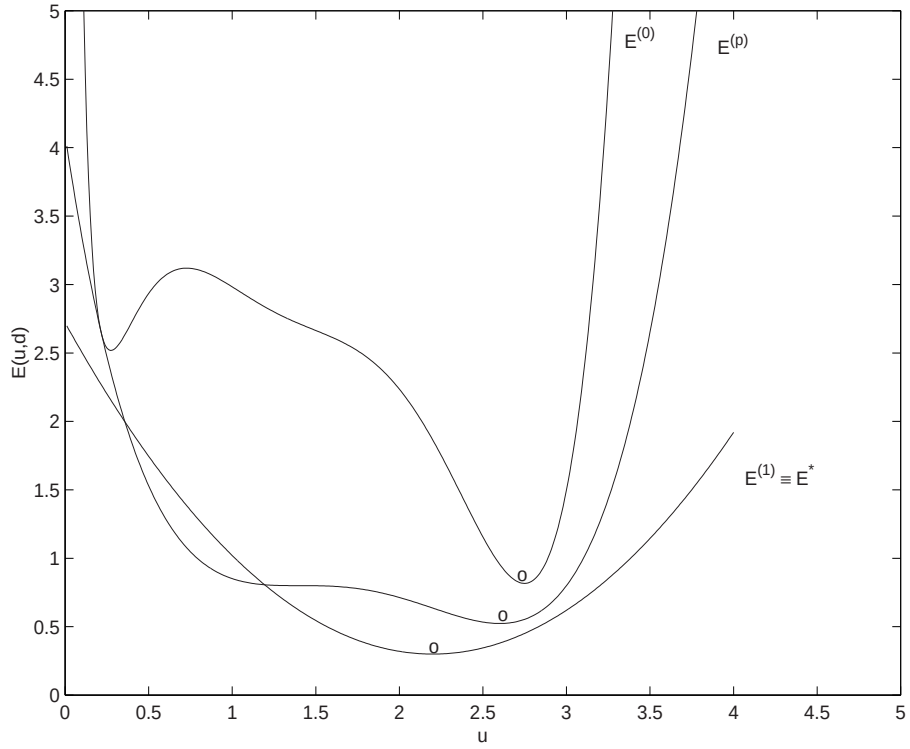

Figura 2.1: Il secondo passo modifica il funzionale da E^* convesso a $E^{(0)} \equiv E$ che è il funzionale dell'energia iniziale da minimizzare.

dell'energia nel caso 1 dimensionale espresso come nella

$$E(u,d) = \int_\Omega ((u(s) - d(s))^2 \, ds + \int_\Omega g(u'(s)) \, ds. \tag{2.3}$$

è che

$$\frac{d^2 g}{dx^2}(x) > -\frac{1}{2}, \tag{2.4}$$

per maggiori dettagli si veda l'appendice. Nel caso del modello della weak membrane[5] la funzione di regolarizzazione che si usa nel funzionale dell'energia iniziale (2.2) è quella quadratica troncata bidimensionale

$$g(\nabla u) = \begin{cases} \lambda^2 |\nabla u|^2 & \text{se } |\nabla \text{u}| < \text{T}/\lambda \\ T^2 & altrimenti \end{cases}$$

che non soddisfa la condizione sufficiente per la convessità (vedi appendice teorema (A.2)) e che quindi deve essere sostituita nel primo passo da un funzionale convesso. Analogamente nel caso del modello 1-dimensionale della weak string la g di partenza del funzionale dell'energia (2.3) è la quadratica troncata 1-dimensionale

$$g(x) = \begin{cases} \lambda^2 x^2 & \text{se } |\text{x}| < \frac{\sqrt{\alpha}}{\lambda} \\ \alpha & \text{altrimenti} \end{cases} \tag{2.5}$$

che non ha la derivata seconda limitata per $|x| = \frac{\sqrt{\alpha}}{\lambda}$ mentre altrove la sua derivata seconda è maggiore uguale a zero. Allora si deve cercare di arrotondare lo spigolo (Figura 2.2) approssimando g con una g^* che verifichi ovunque la condizione (2.4).

Gli algoritmi GNC secondo B&Z sfruttano i modelli di weak continuity e verranno analizzati più in avanti. Prima di tali algoritmi, i metodi usati per costruire uno spazio convesso, facevano riferimento a MFA (Mean Field Annealing) per rilassare e rendere convessa l'Hamiltoniana e all'uso di filtri.

[5] I modelli di weak continuity: weak string e weak membrane, sono introdotti nei successivi capitoli.

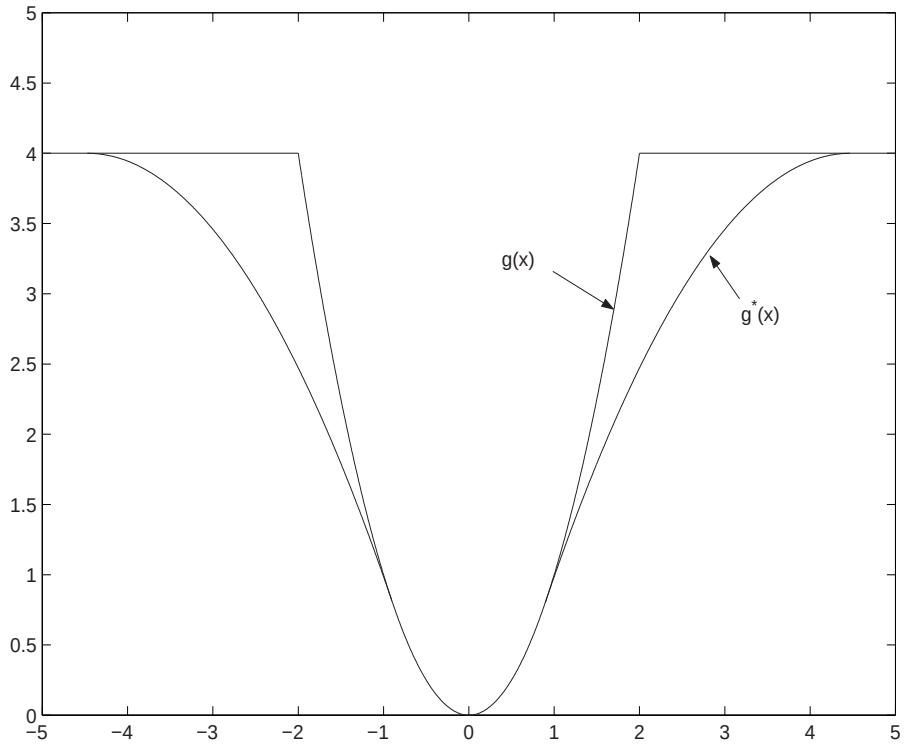

Figura 2.2: La funzione g^* è una approssimazione di g che soddisfa la (2.4) e che rende E^* convesso nel modello della weak string.

2.3 Mean Field Annealing

In questa tecnica si approssima la funzione di regolarizzazione con quella MF-approssimata nella quale si suppone che gli stati (le posizioni) possibili che può assumere la nostra membrana abbiano la distribuzione di Gibbs che soddisfa

$$P(u) = \frac{e^{-E(u,d)/\sigma}}{Z} \tag{2.6}$$

dove $E(u,d)$ rappresenta l'energia (2.1) e Z è una costante di normalizzazione. Allora lo spazio delle soluzioni viene costruito in questa approssimazione attorno allo stato medio della distribuzione. La f che

vien fuori da questa costruzione si chiama *funzione di regolarizzazione MF-approssimata*. La distribuzione di Gibbs (2.6) produce stadi medi determinabili per σ alta mentre per $\sigma = 0$ tutti gli stati sono equiprobabili. Quindi il metodo è costruibile solo per σ alta. Si osserva che calcolare la media di una distribuzione di Gibbs risulta complicato sia per il costo computazionale che per le dipendenze dalle proprietà correnti del campo d. Inoltre vale che la MF-approssimata f_{σ^*}, con σ^* opportunamente alta, porta alla convessità dello spazio e quindi si può utilizzare un algoritmo GNC, nel quale, si parte da σ^* e si ritorna alla sigma iniziale bassa che corrisponde alla funzione di regolarizzazione iniziale f. Se utilizziamo un modello di weak continuità la f iniziale (con $\sigma = 0$) è quella quadratica troncata nel caso 2D

$$f(x) = f_0(x) = \begin{cases} T^2 & se\ |x|^2 > T^2/\lambda \\ \lambda x^2 & se\ |x|^2 < T^2/\lambda \end{cases}$$

mentre quella MF-approssimata generica è

$$f_\sigma(x) = T^2 - \sigma \log(1 + e^{\frac{T^2 - \lambda|x|^2}{\sigma}}).$$

Allora partendo da una σ^* opportunamente alta (tale da rendere il problema convesso), si determina la u_{σ^*} ottima e successivamente si fa decrescere in maniera continua σ determinando la u_σ ottima di ogni passo, fino a raggiungere $\sigma = 0$. In questo procedimento è molto importante scegliere la σ^* iniziale per non cadere in una concavità di un minimo non assoluto.

2.4 Uso dei filtri

Ulteriori procedimenti, che rientrano negli algoritmi GNC, si basano sull'uso dei filtri come strumento per attirare la convessità del funzionale dell'energia. Descriviamo ora il cosiddetto *Smoothness Focusing*

[3]. Esso riesce ad ottenere un'espressione convessa dell'energia sostituendo la funzione di regolarizzazione di partenza con una convoluzione fra la stessa funzione e una gaussiana uniforme di deviazione standard sufficientemente grande

$$g^* = g \otimes G_\sigma.$$

Allora usando un semplice metodo di discesa del gradiente si riesce a minimizzare la E^*. Al decrescere della deviazione standard σ verso 0 ($\sigma_p \to \sigma$) si individua una successione di funzioni di regolarizzazione

$$g^{(p)} = g \otimes G_{\sigma_p}$$

che si avvicinano alla g di partenza e che trasformano E in un funzionale $E^{(p)}$ che ha il minimo sempre più vicino a quello del funzionale di partenza E.
Un procedimento simile che sfrutta la convoluzione gaussiana può essere applicato sulla distribuzione nota a priori degli stati possibili della membrana $P(\nabla_u)$. Se supponiamo che le osservazioni delle ∇_u sono indipendenti fra di loro allora l'interpretazione MAP della funzione di regolarizzazione porta alla

$$g(\nabla_u) = -\log(P(\nabla_u))$$

dove P è la funzione di densità di probabilità delle ∇u. Se ritorniamo al caso 1-dimensionale la convoluzione della funzione di densità di probabilità $P(u)$ con la gaussiana G_σ porta alla funzione di regolarizzazione

$$g_\sigma(u) = -\log(P \otimes G_\sigma)$$

che avrà le derivata seconda positiva (e quindi maggiore di $-\frac{1}{2}$) se la deviazione standard σ della gaussiana è maggiore di quella della distribuzione delle u. Allora con tale sigma si parte da una g_σ che individua un funzionale dell'energia convesso e quindi successivamente

si può applicare GNC facendo decrescere la σ fino a zero per ritornare alla $f(u) = -log(P(u)$ iniziale. Questo procedimento viene detto *Probability Focusing* [3].

Capitolo 3

Weak Continuity

Dati i segnali $d(x)$ della superfice reale $\tilde{u}$, l'obiettivo della restrizione alla *weak continuity* (continuità debole) di una ricostruzione $u(x)$ è quello di determinare i punti dove sono presenti le discontinuità e dove invece la superfice è continua. In particolare una funzione $u(x)$ con continuità debole è una funzione continua a tratti con salti corrispondenti ai punti dove le discontinuità di $d(x)$ sono più rilevanti (Figura 3.1). La restrizione alla weak continuity corrisponde a dei vincoli che nel caso in cui non vengono soddisfatti portano a delle discontinuità, per questo la continuità fragile viene anche detta elastica. Iniziamo con l'esaminare il caso 1-dimensionale dell'Hamiltoniana (2.3); tale modello con la restrizione della weak continuity si dirà modello della *weak string*. In questo funzionale la funzione di regolarizzazione è quadratica

$$g(x) = \lambda^2 x^2.$$

Per introdurre le restrizioni di weak continuity, all'Hamiltoniana

$$D + S = \int_\Omega ((u(x) - d(x))^2 \, dx + \lambda^2 \int_\Omega u'(x)^2 \, dx,$$

si aggiunge una nuova componente. Infatti il termine S forza la continuità mentre l'obiettivo è quello di costruire un modello che preveda la

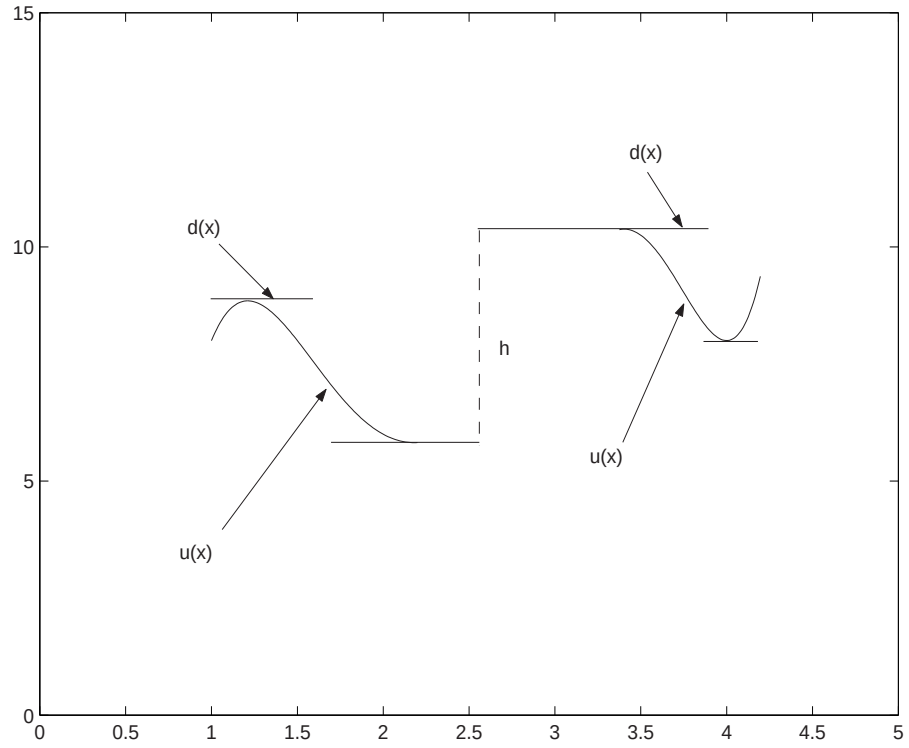

Figura 3.1: La $u(x)$ con la restrizione di weak continuity. Quando il salto h nei dati $d(x)$ è alto allora anche la $u(x)$ presenta un salto nello stesso punto di discontinuità

comparsa di discontinuità nelle ricostruzioni. Perciò all'espressione di sopra si aggiunge un termine P che porta a delle penalità di peso α in particolari punti, ostacolando in essi la ricostruzione continua. Allora il modello della weak string è

$$E(u,d) = D+S+P = \int_\Omega ((u(x)-d(x))^2\,dx+\lambda^2 \int_\Omega u'(x)^2\,dx+P. \quad (3.1)$$

Il parametro λ è detto *di regolarizzazione* infatti più è grande e più la u che vien fuori dalla minimizzazione dell'energia è regolare (cioè continua con derivate basse). Questo parametro è detto anche *dimensione*

di lunghezza o *scala*. Esso regola il bilanciamento dei due termini D e S nell'Hamiltoniana, infatti se è grande viene forzato l'impegno nel rendere minimo S, mentre se è piccolo viene minimizzato maggiormente D, rendendo la ricostruzione più vicina ai dati. Inoltre è una lunghezza caratteristica nella lettura dei dati, infatti permette la classificazione delle discontinuità come isolate o meno. Il termine S viene anche detto *energia elastica della stringa*. Il parametro α, come vedremo in seguito, contribuisce nella ricostruzione, con una penalità (in un certo senso negativa) nei punti in cui è segnalata la discontinuità dei dati $d(x)$. Si osserverà che più α è grande e minore è la probabilità di riscontrare effettivamente una discontinuità, nel punto segnalato, al momento della ricostruzione.
Nel capitolo precedente abbiamo visto che se l'energia è del tipo $D + S$ (2.3) la convessità è garantita dalla scelta di una funzione di regolarizzazione g che soddisfi la (2.4); questa condizione è immediatamente soddisfatta dal funzionale (3.1) del tipo weak string $D + S$, cioé senza il termine P, che quindi è convesso. Mentre, con l'aggiunta del termine di penalità P, si hanno più soluzioni stazionarie (più minimi locali di E), alcune più continue ed altre con più discontinuità e la scelta giusta dipenderà dai parametri α e λ e dai salti h nei punti di discontinuità.

3.1 Introduzione dei line process

Se i nostri dati $d(x)$ presentano dei grossi errori, detti *outlier*, dovuti ad un rumore o a errori nella rilevazione dei segnali, e delle discontinuità effettive (cioè presenti nella soluzione reale), allora l'introduzione del termine P permette l'individuazione delle discontinuità effettive nella ricostruzione finale. Un'espressione possibile del termine di penalizzazione è quella legata all'introduzione di un processo lineare (*line process*) $l(x)$ che varia in 0 e 1 e che segnala la presenza di discontinuità

per $l(x) \to 0$ e l'assenza per $l(x) \to 1$ nella ricostruzione. Quindi per condurre ad una ricostruzione che tenga conto delle informazioni deducibili dai valori di $l(x)$ si deve far comparire il termine di penalità P e scomparire il termine S nei punti in cui $l(x) = 0$, mentre l'opposto deve avvenire nei punti in cui la ricostruzione è continua ($l(x) = 1$). Discretizziamo ora il dominio Ω lineare in N pixel $i = 1, \ldots, N$ e supponiamo che tutti i punti di discontinuità effettiva o outliers appartengono a questo insieme. Allora, indicato con l_i il valore che il line process assume nel pixel i e considerata la stima $u'(i) \simeq u_i - u_{i-1}$, si ottiene l'espressione della weak string nel discreto:

$$E = D+S+P = \sum_{i=1}^{N}(u_i-d_i)^2+\lambda^2\sum_{i=2}^{N}(u_i-u_{i-1})^2\cdot l_i+\alpha\sum_{i=1}^{N}(1-l_i). \quad (3.2)$$

dove $u_i \equiv u(i)$ e $d_i \equiv d(i)$. Come si osserva ora il funzionale dipende dalle d_i e u_i ma anche dai line process $l_i \in [0, 1]$ che possono essere ridotte a semplici variabili booleane. Per cui il problema di minimizzare l'energia diventa quello di determinare le l_i e le u_i che conducono al

$$min_{(u_i,l_i)}E,$$

in cui le d_i sono assegnate. La presenza delle variabili booleane rende indubbiamente più complesso il problema di minimizzare l'energia ed ha anche lo svantaggio di non permettere l'utilizzo degli algoritmi GNC che è indispensabile visto che ora il funzionale non è più convesso.

Capitolo 4

Line process e Regolarizzazione robusta

Nel capitolo precedente abbiamo visto l'importanza da un lato, dell'introduzione dei line process, e la difficoltà dall'altra, di utilizzare un espressione dell'energia con tali variabili booleane per ricostruire l'immagine di partenza. Questa situazione si presenta sia nel modello della weak string (3.2) che in quello della weak membrane. Allora l'ideale sarebbe quello di conservare il vantaggio di costruire immagini con salti di discontinuità che viene garantita dai line process, ma anche evitare che essi compaiano nell'espressione dell'energia, per poter applicare un algoritmo GNC! Per risolvere questo problema ci viene in aiuto la regolarizzazione robusta, che risulta equivalente all'introduzione dei line processes nell'Hamiltoniana. Qui di seguito viene presentata la regolarizzazione robusta, poi vengono definiti in maniera generale i line process ed infine esaminata l'unificazione dei due procedimenti secondo i dettagli di Black e Rangarayan [4].

4.1 Robust regularization

La regolarizzazione forte (robusta) non è altro che l'applicazione della statistica robusta nel problema di regolarizzare certi dati $d(x)$ a disposizione. Il campo della statistica robusta è stato concepito nei problemi di approssimazione quando i dati contengono grossi errori, gli outlier, che non sono conformi con le assunzioni statistiche valide per la maggior parte dei dati (come per esempio sono dei dati che non dipendono dal rumore gaussiano). I principali obiettivi della statistica robusta sono:

- descrivere le strutture migliori che si adattano alla massa dei dati a disposizione;
- determinare i punti con gli outlier e le sottostrutture dei dati deviate.

Le statistiche robuste risolvono allora, in termini analitici, il problema di determinare i parametri $a = (a_1, a_2, \ldots, a_n)$ del modello delle ricostruzioni $u(x, a)$ che meglio si adattano ai dati $d(x)$ a disposizione cioé tali da determinare il

$$\min_a \sum_s \rho(d_s - u(s, a)),$$

dove ρ è la norma dell'errore. Chiaramente la scelta delle differenti ρ-norme porta a differenti stimatori robusti dei dati che avranno una differente sensibilità agli outlier. Ad esempio l'uso dei minimi quadrati ($\rho(x) = x^2$) è molto sensibile agli outlier, perchè tale norma ha un tasso di crescita lineare e ciò non consente il riconoscimento degli outlier. Le norme che prevedono la presenza degli outlier vengono dette *norme robuste*, quando queste norme compaiono anche nel termine di regolarizzazione S dell'energia allora si parla più propriamente di *regolarizzazione robusta.*

Per analizzare il comportamento di una norma robusta ρ, in funzione degli errori (che saranno degli outlier, se $u_s - d_s$ è grande) è indispensabile analizzare la *funzione di influenza* $\phi(x) = \rho'(x)$ che deve evitare una crescita lineare per errori $u_s - d_s$ grandi, altrimenti la norma ρ che comparirebbe nel termine D del funzionale dell'energia (vedi paragrafo successivo) sarebbe troppo grande e forzerebbe l'eliminazione del termine di regolarità S, portando a ricostruzioni simili ai dati e quindi errate. Una classe di norme fortemente robuste sono quelle aventi la funzione d'influenza decrescente all'esterno di un intorno dell'origine, come quelle lorentziane

$$\rho(x, \sigma) = \log\left(1 + \frac{1}{2}\left(\frac{x}{\sigma}\right)^2\right),$$

$$\phi(x) = \frac{2x}{2\sigma^2 + x^2}$$

che hanno dunque un andamento logaritmico con una funzione di influenza (figura 4.1) che decresce per $|x| > \sqrt{2}\sigma$ e quindi gli outlier sono i punti in cui l'errore dei segnali $u_s - d_s$ supera $\sqrt{2}\sigma$. Un'altra classe di norme robuste, che è quella usata nell'algoritmo GNC secondo B&Z, sono le quadratiche troncate

$$\rho(x, \alpha) = \begin{cases} x^2 & se\ |x| < \sqrt{\alpha} \\ \alpha & altrimenti \end{cases} \tag{4.1}$$

cioé del tipo (2.5) che ha la funzione di influenza

$$\phi(x, \alpha) = \begin{cases} 2x & se\ |x| < \sqrt{\alpha} \\ 0 & altrimenti \end{cases}$$

che è nulla per errori rilevanti.

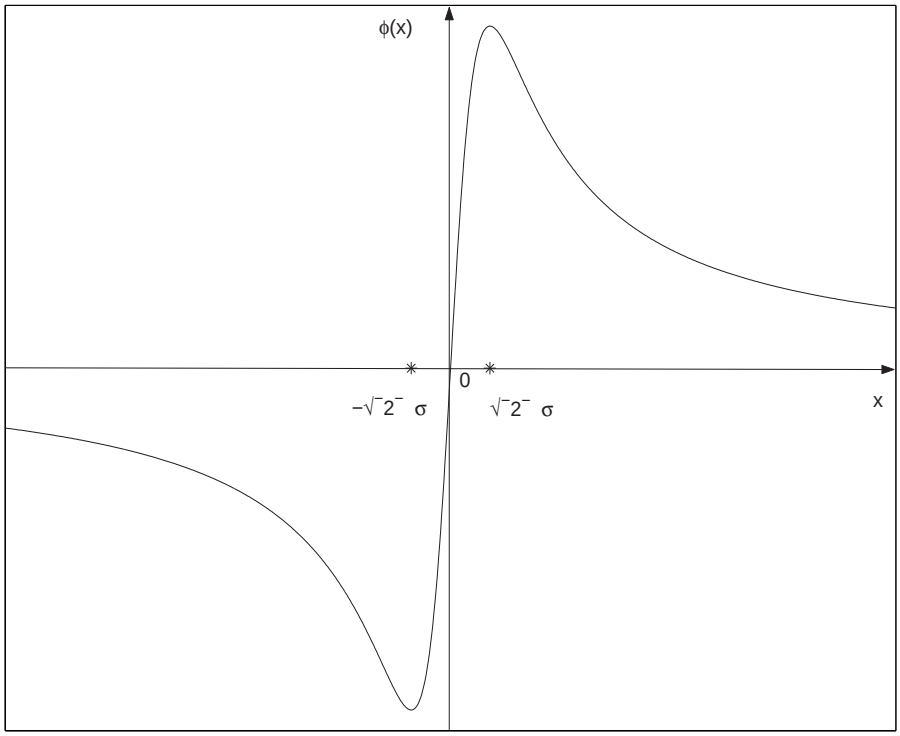

Figura 4.1: Funzione di influenza corrispondente alla norma lorentziana $\rho(x, \sigma)$. Si noti che gli outlier vengono riconosciuti oltre $\sqrt{2}\sigma$, perchè lì la funzione decresce.

4.1.1 Applicazione delle norme robuste nell'Hamiltoniana

L'espressione dell'energia $D + S$ (senza weak continuity) di una stringa $u(x)$ che si adatta ai segnali $d(x)$ nel caso discretizzato

$$E(u, d) = \sum_{i=1}^{N}(u_i - d_i)^2 + \lambda^2 \sum_{i=2}^{N}(u_i - u_{i-1})^2 \qquad (4.2)$$

con le stesse convenzioni di (3.2), è del tipo

$$E(u,d) = \sum_{i=1}^{N} \rho_D(u_i - d_i, \sigma_D) + \lambda^2 \sum_{i=2}^{N} \rho_S(u_i - u_{i-1}, \sigma_S) \tag{4.3}$$

con $\rho_D(x,\sigma) = \rho_S(x,\sigma) = x^2$. Con queste norme non ci troviamo nel caso di regolarizzazione robusta. Infatti se da un lato sul termine D, alla presenza di outlier sui dati $d(x)$, si opera su una struttura deviata nella ricostruzione della stringa, come abbiamo detto precedentemente, dall'altro la norma quadratica sul termine S tenta di rimuovere tutte le irregolarità della funzione (di discontinuità o con derivate elevate) sia se dovute dal rumore che, soprattutto, se dovute ad un salto dell'intensità della stringa reale che è più rilevante nel termine S.

Nella regolarizzazione robusta invece i termini D ed S sono opportunamente pesati mediante delle norme robuste ρ_D e ρ_S che permettono rispettivamente di:

- individuare gli outlier come quei dati che corrispondono agli errori $u_i - d_i$ abbastanza grandi e di dargli un peso piccolo nel termine D;

- trascurare le discontinuità rilevanti $u_i - u_{i-1}$ nel termine S e far pesare nell'Hamiltoniana (4.3) solo le irregolarità minori (dovute dal rumore) portando quindi a regolarizzare la ricostruzione finale solo su questi ultimi punti.

Una volta stabilito la necessità delle norme robuste ρ è necessario sceglierle opportunamente a seconda delle loro proprietà di rendere o meno convessa l'Hamiltoniana (4.3).

I parametri di scala σ delle norme robuste se abbastanza grandi, hanno proprio l'utilità di rendere il problema convesso, ed anche di poter poi utilizzare un algoritmo GNC al diminuire del parametro σ per ritornare al caso iniziale.

L'algoritmo GNC inizia allora con una approssimazione convessa (con σ alto), che porta ad una ρ quadratica (come nel caso della norma lorentziana per $\sigma \to \infty$, [5]) e quindi alla mancata segnalazione di alcun outlier, mentre, al diminuire di σ, la funzione di influenza cresce sempre meno all'esterno di un intorno dell'origine permettendo allora la comparsa di sempre più outlier sui dati e sulla regolarità, per cui la ricostruzione $u(x)$ si adatta a $d(x)$ in maniera sempre più intelligente riconoscendo gli addendi da non considerare nelle sommatorie D e S.

Si osserva che sebbene questo procedimento opera bene, non è garantita la convergenza analitica di $u_\sigma(x)$ alla ricostruzione $u(x)$ con la σ ottima iniziale.

Inoltre può essere provato che l'uso della norma robusta sul termine S porta ad un ottima ricostruzione finale della stringa (o nel caso generale della membrana e quindi dell'immagine) se essa è caratterizzata da intensità costanti a tratti (si veda [5]), che è una delle caratteristiche delle immagini.

4.2 Line Process

Nel capitolo precedente abbiamo introdotto nel caso discreto il termine P con i line process nell'espressione dell'Hamiltoniana (3.2); ora generalizziamo la definizione di line process.

Il line process è un processo lineare $l(x)$ che varia fra 0 e 1 e che permette di ricostruire superfici regolari a tratti, nel senso che forza una ricostruzione continua nei punti dove $l(x) \to 1$ e la presenza di discontinuità nei punti dove $l(x) \to 0$.

L'energia della ricostruzione con un modello generale di line process

su una stringa nel caso discretizzato è allora del tipo

$$E(u,d,l) = D + S + P = \sum_{i=1}^{N}(u_i - d_i)^2 + \lambda^2 \sum_{i=2}^{N}(u_i - u_{i-1})^2 l_i + \sum_{i=1}^{N}\psi(l_i), \tag{4.4}$$

e quindi dipende anche dalla scelta del line process $l(x)$. La funzione $\psi(z)$, detta di *penalità*, deve tendere a zero per $z \to 1$ facendo così scomparire il termine di penalità P (quando $l_i \simeq 1$) e deve tendere ad α, la penalità per una discontinuità nell'Hamiltoniana, per $z \to 0$. Un esempio di funzione di penalità è allora

$$\psi(z) = \alpha(1 - z).$$

La presenza dei line process l_i nella formula (4.4) tiene conto delle possibili violazioni delle restrizioni di regolarità ma non della possibile presenza di segnali outlier nel termine D. Allora per prevedere questa possibilità nel termine di approssimazione dei segnali D si introducono dei line process $m(x)$ ed un'altra funzione di penalità $\psi_D(z)$. Quindi il funzionale dell'energia nel caso discreto diventa

$$E(u,d,l,m) = \sum_{i=1}^{N}[(u_i - d_i)^2 m_i + \psi_D(m_i)] + \lambda^2 \sum_{i=1}^{N}[(u_i - u_{i-1})^2 l_i + \psi_S(l_i)], \tag{4.5}$$

dove si pone $u_0 \equiv u_1$ e $\psi_S(z) = \frac{1}{\lambda^2}\psi(z)$. Il line process $m(x)$ individuerà allora degli outlier nei punti dove tende a zero e delle approssimazioni vicine ai segnali $d(x)$ dove si avvicina all'unità. La formula precedente dell'energia individua il cosiddetto *processo outlier*.

4.3 Equivalenza

Come abbiamo visto, i line process permettono la presenza delle discontinuità nella ricostruzione e prevedono quella degli outlier così com'è

nello spirito del modello con regolarizzazione robusta (4.3), quindi risulta ragionevole il tentativo di provare l'equivalenza fra i due procedimenti. Per far ciò verifichiamo com'è possibile passare da un procedimento ad un altro. Il passaggio dal processo outlier $E(u,d,l,m)$ (4.5) alla formula con la regolarizzazione robusta viene descritto dettagliatamente mentre, per il passaggio inverso, ci si può basare su uno schema presente nell'appendice A.3 e provato su [4].

Il problema di minimizzazione dell'energia nel caso del process outlier è

$$\min_{u,m,l} \sum_{i=1}^{N} [(u_i - d_i)^2 m_i + \psi_D(m_i)] + \lambda^2 \sum_{i=1}^{N} [(u_i - u_{i-1})^2 l_i + \psi_S(l_i)]$$

che risulta equivalente a determinare il

$$\min_{u} \left(\min_{m} \sum_{i=1}^{N} [(u_i - d_i)^2 m_i + \psi_D(m_i)] + \min_{l} \lambda^2 \sum_{i=1}^{N} [(u_i - u_{i-1})^2 l_i + \psi_S(l_i)] \right). \tag{4.6}$$

Costruiamo ora le ρ norme :

$$\rho_D(x) = \inf_{0 \le z \le 1} x^2 z + \psi_D(z,)$$

$$\rho_S(x) = \inf_{0 \le z \le 1} x^2 z + \psi_S(z),$$

da cui sostituendole nella (4.6) otteniamo il modello con la regolarizzazione robusta

$$\min_{u} \sum_{i=1}^{N} \rho_D(u_i - d_i) + \lambda^2 \sum_{i=1}^{N} \rho_S(u_i - u_{i-1}).$$

Questa equivalenza ed in particolare il passaggio da outlier process ad un modello con regolarizzazione robusta permette un costo minore nonché la possibilità di applicare algoritmi GNC sulle norme robuste.

Un primo esempio di equivalenza sussiste fra le norme lorentziane e la funzione di penalità del tipo

$$\psi(z) = const(z - 1 - \log(z))$$

oppure fra la norma quadratica troncata $\rho(x, \alpha)$ espressa in (4.1) e la funzione di penalità

$$\psi(z) = \alpha(1 - z).$$

Infatti

$$\inf_z x^2 z + \alpha(1 - z) = \inf_{0 \le z \le 1} (x^2 - \alpha)z + \alpha = \begin{cases} x^2 & se \quad x^2 < \alpha \\ \alpha & se \quad x^2 \ge \alpha \end{cases}$$

da cui per la definizione

$$\rho(x) = \inf_{0 \le z \le 1} x^2 z + \psi(z)$$

si ha

$$\rho(x) = \begin{cases} x^2 & se \quad |x| < \sqrt{\alpha} \\ \alpha & se \quad |x| \ge \sqrt{\alpha} \end{cases} . \tag{4.7}$$

Capitolo 5

Uso dei parametri α e λ

In questo capitolo esamineremo gli effetti che le variazioni dei parametri α e λ producono all'interno del modello della weak string del tipo (3.1), e successivamente questi risultati verranno generalizzati nel modello della weak membrane che verrà ivi introdotto. Per semplificare i calcoli si considera P del tipo $P = \alpha Z = \alpha \int 1 - l(x)dx$ e quindi l'Hamiltoniana nel modello della weak string diventa

$$E(u,d) = D+S+P = \int_\Omega ((u(x)-d(x))^2\, dx + \lambda^2 \int_\Omega u'(x)^2\, dx + \alpha Z. \quad (5.1)$$

Come abbiamo detto nel capitolo 3 una caratteristica fondamentale della weak continuity è l'elasticità, cioé la capacità di rilevare discontinuità nei dati solo per dei salti rilevanti; il prossimo paragrafo ci permette di stabilire il valore limite di salto nei segnali, che porta all'individuazione della discontinuità nella ricostruzione. Questo salto limite viene detto *limite di sensitività* e indicato con h_0.

5.1 Limite di sensitività

Supponiamo di considerare l'intervallo $[0, L]$ su cui viene rilevato il segnale $d(x)$ regolare (senza discontinuità) allora si può provare (Appen-

dice A.4) che l'energia nel modello continuo

$$E(u,d) = \int_0^L ((u(x) - d(x))^2 \, dx + \lambda^2 \int_0^L u'(x)^2 \, dx, \qquad (5.2)$$

assume il valore minimo

$$E(\overline{u}) = \int_0^L d(x)(\overline{u}(x) - d(x))dx,$$

con chiaramente $\overline{u}$ la migliore ricostruzione continua su $[0, L]$.

Se consideriamo lo spazio delle stringhe continue in [0,L] che soddisfano il vincolo $u(0) = z$ (Figura 5.1) con $z \neq \overline{z} \equiv \overline{u}(0)$, otterremo che la ricostruzione $u(x)$ di tale spazio, che minimizza l'energia, porta ad un incremento di energia $\Delta E = E(u) - E(\overline{u})$, che è proporzionale al quadrato della differenza $u(0) - \overline{u}(0)$, cioè vale

$$\Delta E = \epsilon(z - \overline{z})^2.$$

Nel caso $L >> \lambda$ si prova che $\epsilon \simeq \lambda$ per cui

$$\Delta E = \lambda(z - \overline{z})^2,$$

che vale anche se $L = +\infty$.

Ora supponiamo di avere dei segnali in $[-L_2, L_1]$ con una discontinuità in 0. In questa situazione si vuole valutare l'incremento dell'energia prodotto da una ricostruzione continua in 0 rispetto a quella minima nel modello (5.2) in $[-L_2, L_1]$. Per far ciò possiamo sfruttare il risultato precedente perchè imporre la continuità in 0 è equivalente a fissare il passaggio per un punto $(0, z)$. Definiamo la ricostruzione ottima $\overline{u}$ come quella definita in tutto $[-L_2, L_1]$ (Figura 5.1), che minimizza l'energia nel modello continuo (5.2):

- in $[0, L_1]$ (dove si pone $\overline{u}_+ \equiv \overline{u}$),
- e in $[-L_2, 0]$ (dove $\overline{u}_- \equiv \overline{u}$).

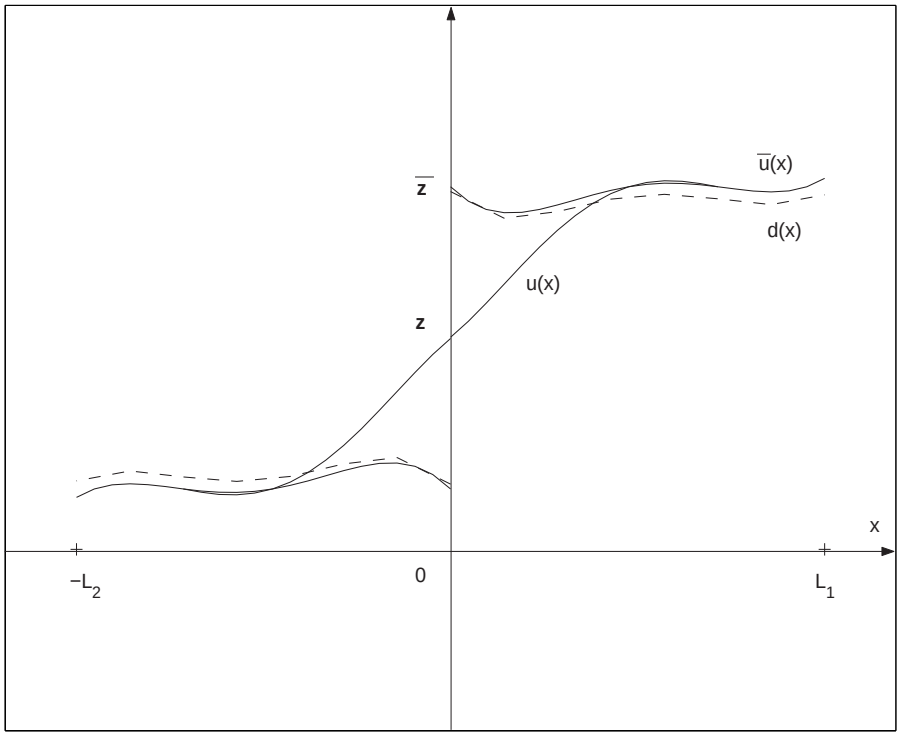

Figura 5.1: La linea tratteggiata rappresenta il segnale $d(x)$ in tutto $[-L_2, L_1]$, mentre $\overline{u}$ è la ricostruzione ottima per il modello $D+S$. $u(x)$ è la stringa che soddisfa il vincolo $u(0) = z$.

Questa ricostruzione è proprio quella che minimizza l'energia in tutto $[-L_2, L_1]$.

Allora l'incremento dell'energia rispetto a $\overline{u}$, ottenuto nell'imporre la continuità (cioè $u(0) = z$), è proporzionale alla somma delle differenze $\overline{u}_+(0) - u(0)$ e $u(0) - \overline{u}_-(0)$, per cui ΔE è proporzionale al salto h di discontinuità di $d(x)$ e vale

$$\Delta E = E(u) - E(\overline{u}) = h \left(\frac{1}{\epsilon_1} + \frac{1}{\epsilon_2} \right)^{-1}.$$

Nel caso in cui $L_1, L_2 >> \lambda$ si ha $\lambda = \epsilon_1 = \epsilon_2$ e

$$\Delta E = \frac{\lambda h^2}{2}. \tag{5.3}$$

Grazie a quest'ultima espressione di ΔE è possibile determinare h_0.

Consideriamo sempre i dati $d(x)$ con una sola discontinuità in 0, allora, con il modello dell'energia della weak string (5.1), la ricostruzione ottima (che minimizza E) presenta la discontinuità in 0 se la sua energia $E = \min(D+S) + \alpha$ è minore dell'energia di una ricostruzione continua in 0 del tipo $D+S$. Siccome la differenza $D+S-\min(D+S)$ è pari all'incremento dell'energia nel modello (5.2) in $[-L_2, L_1]$, per la (5.3), otteniamo che compare la discontinuità nella ricostruzione ottima se

$$\Delta E > \alpha$$

che è equivalente alla condizione

$$h > \sqrt{\frac{2\alpha}{\lambda}},$$

che permette allora di definire il limite di sensitività:

$$h_0 = \sqrt{\frac{2\alpha}{\lambda}}. \tag{5.4}$$

5.1.1 Discontinuità interagenti

Nell'analisi precedente abbiamo supposto che nell'intervallo $[-L_2, L_1]$ era presente una sola discontinuità, ma nella realtà i segnali presentano parecchie discontinuità. Il caso più semplice di due discontinuità è quello del cappello a cilindro che ha un doppio gradino di uguale altezza h (Figura 5.2), sul quale determineremo l'associato limite di sensitività che verrà poi valutato nel caso generale; proveremo inoltre che il limite di sensitività cambia se i punti di discontinuità sono abbastanza vicini. D'ora in poi indicheremo con a la distanza fra due discontinuità consecutive.

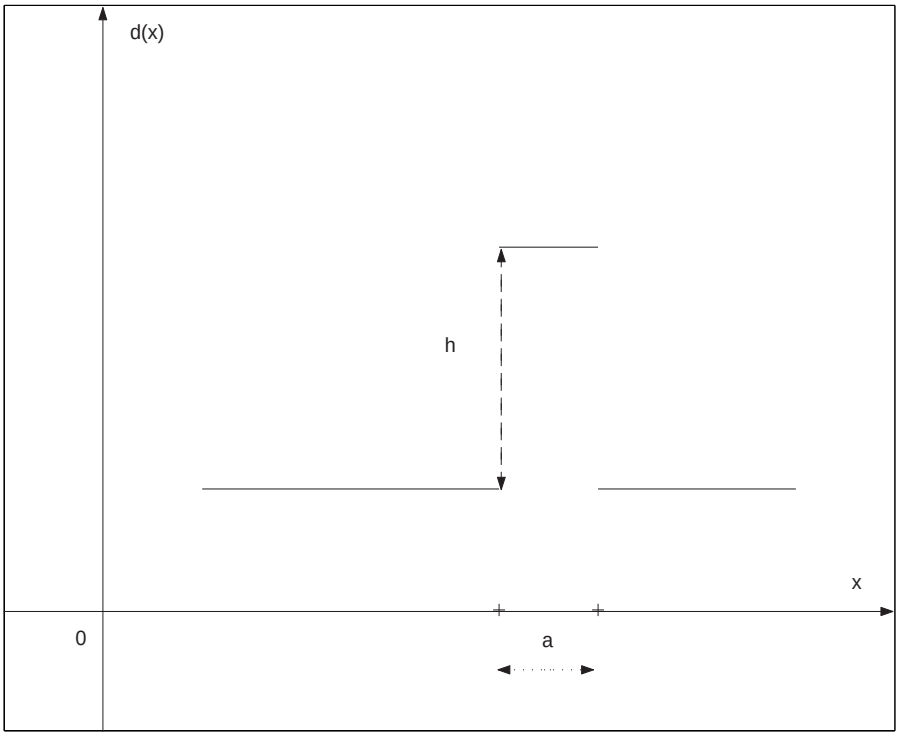

Figura 5.2: Funzione "cappello a cilindro" di altezza h.

Iniziamo con l'osservare che il caso di una sola discontinuità, si può vedere come la situazione di più discontinuità che distano a, con $a > L_1$ e $a > L_2$ (basta allargare l'intervallo di ricostruzione), dove $a >> \lambda$ giacché il risultato precedente è stato ottenuto per $L_1, L_2 >> \lambda$.

Questa analogia porterà allo stesso limite di sensibilità.

Se supponiamo $a >> \lambda$ si ha che l'energia della ricostruzione continua che minimizza E nel modello $D + S$ è, nel caso che la base del cilindro poggia sull'asse delle x (si veda l'appendice di [2])

$$E_c = h^2 \lambda \left(1 - \exp\left(-\frac{a}{\lambda}\right)\right).$$

Mentre la ricostruzione ottima che presenta le due discontinuità ha

l'energia $E_d = 2\alpha$. Per cui nel caso in cui

$$E_d = 2\alpha < E_c = h^2\lambda\left(1 - \exp\left(-\frac{a}{\lambda}\right)\right) \simeq h^2\lambda,$$

ossia se vale

$$h > h_0 = \sqrt{\frac{2\alpha}{\lambda}},$$

avremo la certezza di individuare le due discontinuità con il modello di weak string.

Mentre nel caso inverso $a << \lambda$ siccome $(1 - \exp(\frac{a}{\lambda})) \simeq \frac{a}{\lambda}$ vale che $E_c = h^2 a$ e quindi $E_d = 2\alpha < E_d$ se

$$h > \sqrt{\frac{2\alpha}{a}} = \sqrt{\frac{\lambda}{a}} \cdot h_0$$

Quindi le discontinuità interagiscono nella ricostruzione se $a << \lambda$ portando ad una modifica del limite di sensitività.

I risultati ottenuti possono essere estesi anche al caso più generale in cui le discontinuità interagenti ($a << \lambda$) nei dati sono relative a due salti di altezza diversa h_1, h_2 (Figura 5.3).

In questa situazione con $h_2 > h_1$, sono 3 le possibili configurazioni della stringa che verranno ottenute se le rispettive energie, ottenute minimizzando i relativi modelli d'energia, sono minori rispetto alle altre. L'elenco dei possibili valori dell'energia in ogni configurazione è descritto qui di seguito:

- nessuna discontinuità, $E = \frac{1}{2}\lambda(h_1 - h_2)^2 + ah_1h_2$;
- discontinuità di salto h_2, $E = ah_1^2 + \alpha$;
- entrambe le discontinuità, $E = 2\alpha$.

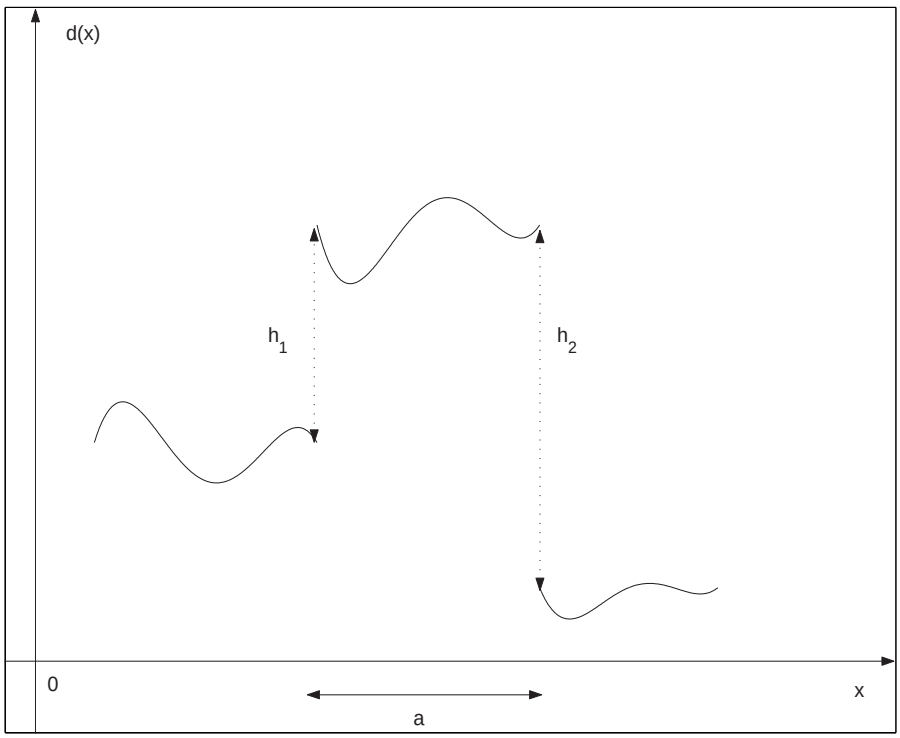

Figura 5.3: Segnali con due salti di altezza diversa h_1, h_2

5.2 Discontinuità spurie generate dalla digitalizzazione

Un aspetto interessante del comportamento del modello con weak continuity è quello legato alla comparsa di discontinuità spurie dovute alla digitalizzazione. Infatti in presenza di una sola discontinuità di salto h, la digitalizzazione mediante dei campioni centrati nella zona di discontinuità (Figura 5.4) porta al campionamento del segnale verso un valore non esistente nei dati, facendo comparire una discontinuità spuria oltre a quelle già presenti.

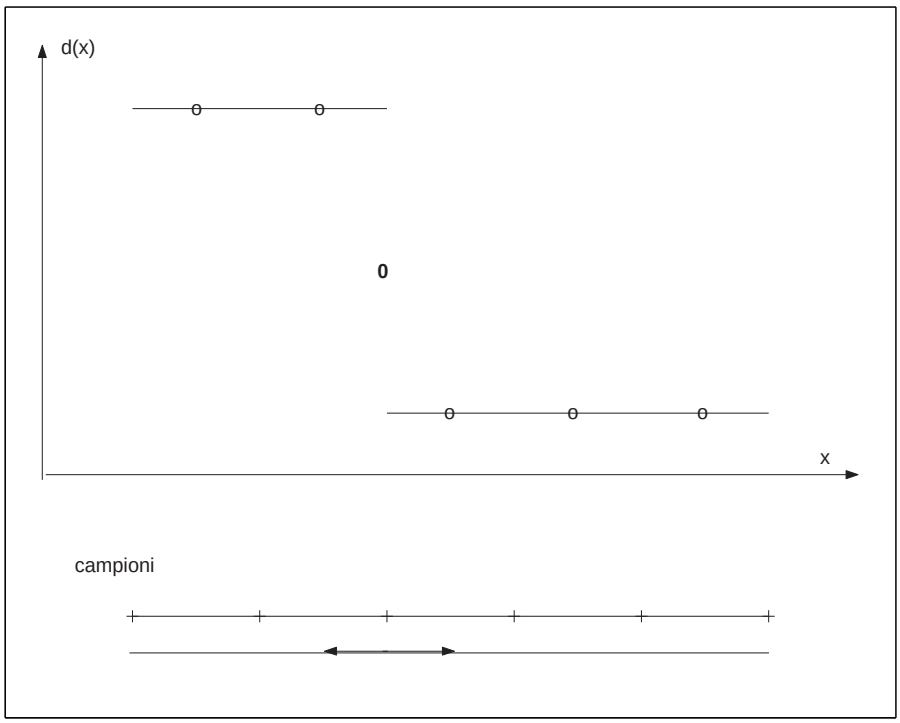

Figura 5.4: Il gradino $d(x)$ viene campionato su intervalli di uguale misura ciò produce dei digit (i cerchietti). Un campionamento centrato nella zona di discontinuità (rappresentato dalle freccette) crea un digit (quello scuro) che corrisponde alla scoperta di una discontinuità spuria.

L'energia di una ricostruzione con la discontinuità spuria è pari a 2α mentre nel caso, della sola discontinuità reale, in analogia al caso generale di $a << \lambda$ visto precedentemente, l'energia è $E = \alpha + \frac{1}{4}h^2a$, come se la discontinuità spuria sia una discontinuità effettiva nei dati. Allora la discontinuità spuria non appare se

$$\alpha + \frac{1}{4}h^2a < 2\alpha$$

ossia, ricordando la definizione (5.4) di h_0, se

$$\frac{h}{h_0} < \sqrt{\frac{2\lambda}{a}}$$

cioè per λ abbastanza grande o per a abbastanza piccolo. Si noti che la distanza a fra le discontinuità nei dati (digitalizzati) approssima l'ampiezza dell'intervallo di campionamento, quindi più il campionamento è preciso e minore è la probabilità che appaia una discontinuità spuria. Tale distanza a essendo minore dell'unità garantisce la ricostruzione senza discontinuità spurie per

$$h < h_0\sqrt{2\lambda}.$$

5.3 Rumore e discontinuità spurie

L'errore indotto dal rumore nella localizzazione delle discontiuità è molto piccolo. Se indichiamo con u_ϵ la ricostruzione dei dati $d(x)$ che presenta la discontinuità di salto h spostata di ϵ piccolo rispetto a quella dei dati allora solo se la deviazione standard del rumore σ verifica

$$\sigma > \frac{h}{2}$$

vale

$$E(u_\epsilon, d) < E(u_0, d),$$

e quindi la discontinuità viene rilevata in punto diverso da quello reale.

Una misura dell'immunità dal rumore nella ricostruzione della stringa, per quanto riguarda la comparsa di discontinuità spurie, è data dal parametro α. Infatti già sappiamo che un α alto, nel modello della weak string, incoraggia la continuità e la regolarità, appiattendo allora il contributo del rumore che rende invece la funzione più irregolare. In generale si può provare che per α che soddisfa la relazione

$$\alpha > 2\sigma^2$$

nella ricostruzione risultante non compaiono delle discontinuità spurie. La dimostrazione di questo risultato viene provato nel caso in cui $d(x) = 0$ qui di seguito.

Indicata ancora con u_ϵ la ricostruzione di $d(x)$ che minimizza l'energia (5.1) fra tutte quelle con una discontinuità (spuria) in $x = \epsilon$, e con $u_c(x)$ la migliore ricostruzione continua (cioè $u_c = 0$), vale l'equazione

$$E(u_\epsilon, d) - E(u_c, d) = \alpha - R^2$$

dove R è una variabile aleatoria gaussiana con deviazione standard $\frac{\sigma}{\sqrt{2}}$ e media 0. L'immunità dal rumore si ottiene allora se $E(u_c, d) < E(u_\epsilon, d)$ cioè se $\alpha > R^2$. Quest'ultima condizione è soddisfatta nel 95% dei casi se $\alpha > 2\sigma^2$, (si veda l'appendice A.5).

5.4 Modello della weak membrane

L'energia nel modello della weak string (5.1) può essere estesa al caso bidimensionale ottenendo il modello della *weak membrane* semplificato

$$E(u, d) = D + S + P = \int_\Omega (u - d)^2 dA + \lambda^2 \int_\Omega |\nabla u|^2 dA + \alpha \int_L dl \quad (5.5)$$

dove $\Omega \subset \mathbb{R}^2$ è una porzione di piano, dA è un elemento infinitesimo di Ω e L è un suo sottoinsieme. L'insieme L racchiude le zone di discontinuità su cui si vuole costruire la superfice $u(x, y)$, quindi $Z = \int_L dl$ è la misura (lunghezza) dell'insieme dei punti di discontinuità. Questo significato di Z è analogo a quello del modello della weak string dove $Z = \int 1 - l(x)dx$ conta proprio il numero delle discontinuità. Per quando riguarda il significato dei due parametri si rimanda al capitolo 3; mentre in questo e nei successivi paragrafi essi verranno usati per descrivere i comportamenti principali della ricostruzione.

Si osserva che, nel modello continuo di membrana $D + S$, la ricostruzione ottima $\overline{u}(x)$ è del tipo

$$\overline{u}(x) = \int_\Omega G(x, x')d(x')dA,$$

con $G(x, x') = \dfrac{1}{2\pi\lambda^2} K_0\left(\dfrac{|x - x'|}{\lambda}\right)$ e K_0 è una funzione di Bessel del secondo tipo (si veda [2]). L'energia minima associata è

$$E_c = E(\overline{u}) = \int_\Omega d(x)(d(x) - \overline{u}(x))dA \tag{5.6}$$

Per valutare il limite di sensitività delle discontinuità h_0 con la presenza di più discontinuità si procede in maniera analoga del caso della weak string, valutando l'energia nel caso di segnali riferiti questa volta a un cappello a cilindro in 2D (Figura 5.5). Si può provare (appendice [2]) che l'energia della ricostruzione continua $\overline{u}$ associata ai segnali $d(x, y)$, nel caso delle discontinuità isolate $\lambda << a$, è per la (5.6)

$$E_c = E(\overline{u}) = \pi a \lambda h^2.$$

Invece nel caso in cui sono presenti le discontinuità sulla circonferenza di raggio a la (5.5) porta al minimo per $u \equiv d$ con

$$E_{disc} = 2\pi\alpha a.$$

Per cui saranno presenti le discontinuità nella ricostruzione se

$$2\pi\alpha a < \pi a \lambda h^2,$$

ossia se il salto h e maggiore del limite di sensitività

$$h_0 = \sqrt{\frac{2\alpha}{\lambda}}.$$

Se invece $a << \lambda$, si ha che la ricostruzione continua migliore ha un'energia indipendente da λ

$$E_c = \pi a^2 h^2$$

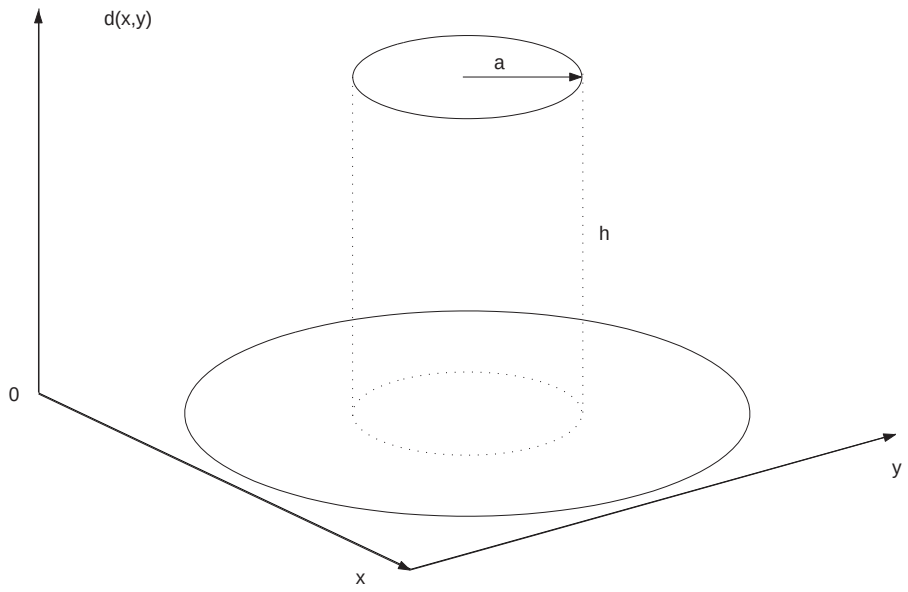

Figura 5.5: Cappello a cilindro $d(x, y)$ di altezza h con il raggio del cerchio superiore pari ad a,

ed avendo ancora nel caso discontinuo $u \equiv d$, compariranno le discontinuità se

$$E_{disc} = 2\pi\alpha a < E_c,$$

ossia se

$$h > \sqrt{\frac{\lambda}{a}}\, h_0.$$

Questi limiti di sensitività, determinati nel caso particolare del cappello a cilindro, hanno le stesse espressioni per una qualsiasi membrana $d(x, y)$.

5.5 Comportamento della ricostruzione

Per quanto visto nel paragrafo precedente il limite di sensitività ha la stessa espressione del caso della weak string. Inoltre si può osservare che, nel modello della weak membrane, anche gli altri comportamenti soddisfano le stesse relazioni del modello weak string.

Il seguente schema riassume tali comportamenti al variare dei parametri α e λ:

- *Scala per le distanza dei punti (x,y)*
  Misura: λ
  Comportamento:
  - se $\lambda << a$ le discontinuità sono isolate, quindi non c'è influenza nella ricostruzione in un intorno di una discontinuità, da parte delle discontinuità vicine.
  - se $\lambda >> a$ le discontnuità sono interagenti (non isolate).
- *Limite di sensitività per le discontinuità isolate* $h_0 = \sqrt{\frac{2\alpha}{\lambda}}$
  Comportamento:
  - appaiono le discontinuità se $h > h_0$
  - ricostruzione continua se $h < h_0$.
- *Limite di sensitività per discontinuità interagenti distanti a* $h'_0 = \sqrt{\frac{\lambda}{a}} h_0$.
- *Resistenza al rumore per la comparsa di discontinuità spurie*
  Misura: α
  Comportamento: non appaiono le discontinuità spurie se $\alpha > 2\sigma^2$ con σ^2 la varianza della distribuzione del rumore.

- *Comparsa di discontinuità spurie per la digitalizzazione*
  La misura è $h_0\sqrt{2\lambda}$
  comportamento: non appaiono le discontinuità spurie se

$$h < h_0\sqrt{2\lambda}. \tag{5.7}$$

5.6 Modello adattivo della weak membrane

Lo schema precedente mette in luce che c'è difficoltà nello scegliere i parametri α e λ che evitino la presenza di discontinuità spurie, mantenendo una certa sensitività controllabile e indipendente dai dati $d(x, y)$.

Per quanto riguarda l'indipendenza si osserva che, un salto h di 100, per dati che variano in $[10000, 100000]$, è cosa ben diversa se analizzato su dati che variano in $[100, 1000]$. Per questo, se si vuole una sensitività nella ricostruzione "relativamente costante", questa deve essere adattata ai dati a disposizione. Un esempio può essere quello di considerare come dati le $d'(x, y) = \log(d(x, y))$, dove le $d(x, y)$ sono i segnali a disposizione. Così avremo che la variazioni sui nuovi dati

$$\Delta d'(x, y) = \frac{\Delta d(x, y)}{d(x, y)}$$

è proporzionale all'intensità dei segnali di partenza. Allora il limite di sensitività su d' lavora in maniera ben pesata sulle variazioni reali di d. Però a questo punto sorge il problema che, a basse intensità $d(x, y)$, le variazioni dovute al rumore che si sommano con quelle effettive, hanno un peso rilevante nel conto finale di $\Delta d'$, portando così a ricostruzioni falsate.

Qui di seguito viene introdotto uno schema adattivo sul modello della weak membrane, che evita i problemi legati al rumore e alla digitalizzazione e che usa un limite $h_0 = h_0(x, y)$ proporzionale ai segnali a disposizione.

Schema

1. Si acquisiscono in input i segnali della membrana misurata $d(x, y)$, il parametro di regolarizzazione (o scala) $\lambda > 0$ e la misura (negativa) della sensitività alle variazioni nei dati $0 < \eta < 1$.

2. Si costruisce una membrana continua del tipo $D + S$, con un parametro λ' relativamente alto per rendere quanto più regolare e continua la membrana. Si sceglie anche un parametro α' abbastanza alto da superare $2\sigma^2$ e tale che $h_{media} < h_0$, in maniera tale da avere rispettivamente immunità dal rumore e una ricostruzione con poche discontinuità.

3. La ricostruzione continua ottenuta al passo precedente la indichiamo con $I_{av}(x, y)$. Infatti, approssimativamente, una regolarizzazione forte porta ad una membrana con intensità costante in un intorno di ogni punto, e di valore pari alla media delle intensità dei segnali iniziali nell'intorno. Si pone allora

$$h_0(x, y) = \eta I_{av}(x, y). \tag{5.8}$$

   Questo limite di sensitività è allora proporzionale punto per punto all'intensità media di un opportuno intorno del punto. Esso sarà più grande se la misura η è più grande cioè se si vuole costruire una membrana meno sensibile alle variazioni nei dati e quindi alle discontinuità. Per questo η si può considerare come una misura negativa della sensitività.

4. Viene applicato un algoritmo GNC sul modello della weak membrane uguale a quello che esamineremo nel prossimo capitolo, ma con la differenza che $\alpha = \alpha(x, y)$ è una funzione con il valore dedotto dalla definizione di $h_0(x, y)$ (5.8), cioé

$$\alpha(x, y) = \lambda \frac{h_0(x, y)^2}{2} = \frac{\lambda \eta^2 I_{av}(x, y)^2}{2}.$$

La scelta dei parametri di input λ e η deve essere tale da evitare anche le discontinuità spurie dovute alla digitalizzazione. Cioè si deve verificare la relazione (5.7) per ogni salto h nei dati. Osserviamo che se, in (x, y), la media del segnale è $I_{av}(x, y)$ allora è lecito attendersi che al più il salto h dei dati iniziali è $2I_{av}(x, y)$, che è ottenuto quando d varia da 0 a $2I_{av}(x, y)$. Quindi per la (5.8) vale sempre

$$h \leq 2I_{av} < 2\frac{h_0(x, y)}{\eta}.$$

Allora basta scegliere η e λ tali che

$$2\frac{h_0(x, y)}{\eta} < h_0\sqrt{2\lambda}$$

cioé

$$\eta \geq \sqrt{\frac{2}{\lambda}}$$

per non avere discontinuità spurie nella ricostruzione finale.

Osservazione operativa 5.1 *Questo schema adattivo viene utilizzato nel file "Gnc.c" allegato in coda alla relazione, con delle modifiche:*

- *Per risparmiare lo spazio in memoria, si è deciso di imporre la dimensione dell'intorno del punto con intensità $I_{a,v}$ costante. Allora l'intensità $I_{a,v}$ varia regione per regione e non punto per punto. Quindi indicata con (x, y) una regione dell'immagine, $I_{av}(x, y)$ è la media dell'intensità in tale regione.*

- *Quindi anche $\alpha(x, y)$, dedotta dalla (5.8), è relativa ad una regione di coordinate (x, y).*

- *Una stima dei salti medi h_{media} è la deviazione standard di tutte le intensità dell'immagine iniziale.*

- *La condizione per l'immunità dal rumore usata è $\alpha \cdot \lambda > 2\sigma^2$ in accordo con la condizione di immunità nell'applicazione dell'algoritmo GNC (prossimo capitolo paragrafo 6.3.2).*

Capitolo 6

Algoritmo GNC di Blake & Zisserman

L'algoritmo GNC redatto da Blake e Zisserman [2] usa un modello di weak continuity discretizzato con i line process che nel caso della weak string è del tipo (3.2). Per l'energia discretizzata espressa in questa forma valgono i risultati del precedente capitolo anche se sono stati determinati per un modello diverso (5.1).

Ora esprimiamo il modello (3.2) nella forma (4.5) che presenterà la funzione di penalità nel solo termine di regolarità:

$$E = \sum_i (u_i - d_i)^2 + \lambda^2 [\sum_i (u_i - u_{i-1})^2 l_i + \psi_S(l_i)]$$

con $\psi_S(z) = \frac{\alpha}{\lambda^2}(1 - z)$.

Come abbiamo già detto nel capitolo 4, per poter applicare un algoritmo GNC su un modello con i line process lo si trasforma nella forma equivalente con le norme robuste.

Tale modello con la norma robusta nel termine di regolarità è

$$\sum_i (u_i - d_i)^2 + \lambda^2 g_s(u_i - u_{i-1}),$$

dove g_s è la norma robusta della forma (4.7), con la sostituzione del parametro α con $\dfrac{\alpha}{\lambda^2}$, cioé

$$g_S(x) = \inf x^2 z + \psi_S(z) = \begin{cases} x^2 & se\ |x| < \frac{\sqrt{\alpha}}{\lambda} \\ \frac{\alpha}{\lambda^2} & se\ |x| \geq \frac{\sqrt{\alpha}}{\lambda} \end{cases} .$$

Allora, posto

$$g(x) \equiv \lambda^2 g(x) = \begin{cases} \lambda^2 x^2 & se\ |x| < \frac{\sqrt{\alpha}}{\lambda} \\ \alpha & se\ |x| \geq \frac{\sqrt{\alpha}}{\lambda} \end{cases}$$

otteniamo il modello dell'energia della weak string secondo B&Z :

$$E = \sum_i (u_i - d_i)^2 + g(u_i - u_{i-1}). \tag{6.1}$$

Questo risultato viene generalizzato nel caso della weak membrane ottenendo il modello

$$E = \sum_i \sum_j (u_{i,j} - d_{i,j})^2 + g(u_{i,j} - u_{i,j-1}, u_{i,j} - u_{i-1,j}), \tag{6.2}$$

con

$$g(x,y) = \begin{cases} \lambda^2(x^2 + y^2) & se \quad ||(x,y)||_2 \equiv \sqrt{x^2 + y^2} < \frac{\sqrt{\alpha}}{\lambda} \\ \alpha & se \quad \|||x,y||_2 \geq \frac{\sqrt{\alpha}}{\lambda} \end{cases} .$$

L'algoritmo GNC applicato a tali modelli è formato da due passi:

1. determinazione del funzionale convesso

   $$E^* = D + S^* = \sum_i (u_i - d_i)^2 + g^*(u_i - u_{i-1})$$

   che approssima il funzionale iniziale;

2. definizione di una sequenza di funzionali $\{E^{(p)}\}$ che, partendo dal funzionale convesso $E^{(1)} \equiv E^*$, porti ad approssimazioni sempre migliori del funzionale $E \equiv E^{(0)}$ fino a raggiungerlo, per p che varia da 1 a 0. In ognuno di questi funzionali si determina il minimo partendo dalla posizione raggiunta nella discesa del precedente funzionale.

6.1 Determinazione del funzionale convesso

Una condizione necessaria per stabilire se il funzionale della weak string

$$E^* = D + \sum_i g^*(u_i - u_{i-1})$$

è convesso è già stata data nel capitolo 2, (2.4); in cui si è accertato che il modello iniziale (6.1) non è convesso. Un'altra condizione sufficiente (più generale) che garantisce la convessità di E^* è

$$\exists c^* > 0 \text{ tale che } \forall \mathrm{x}\ (\mathrm{g}^*)''(\mathrm{x}) \geq -\mathrm{c}^* \text{ e } \mathrm{v}_{\max} \leq \frac{2}{\mathrm{c}^*} \tag{6.3}$$

con v_{max} l'autovalore massimo di un opportuna matrice Q^TQ; per maggiori dettagli si veda l'appendice (teorema (A.3)). Allora notato che $v_{max} = 4$ è semplice costruire un funzionale convesso nel modello weak string. Ad esempio, nel procedimento di B&Z, si sceglie un' opportuna $g^*(x)$ tale che

$$g^*(x) \leq g(x), \quad (g^*)''(x) \geq -c^*,$$

con $c^* = \dfrac{2}{v_{max}} = \dfrac{1}{2}$, così definita

$$g^*_{\alpha,\lambda}(x) = \begin{cases} \lambda^2 x^2 & se \quad |x| < q \\ \alpha - c^* \dfrac{(|x| - r)^2}{2} & se \quad q \leq |x| < r \\ \alpha & se \quad |x| \geq r \end{cases} . \tag{6.4}$$

dove

$$r^2 = \alpha \left(\frac{2}{c^*} + \frac{1}{\lambda^2} \right), \quad e \quad q = \frac{\alpha}{\lambda^2 r} .$$

D'ora in poi questa funzione viene anche indicata con $g^*(x)$, il suo grafico è presente in figura (6.1)

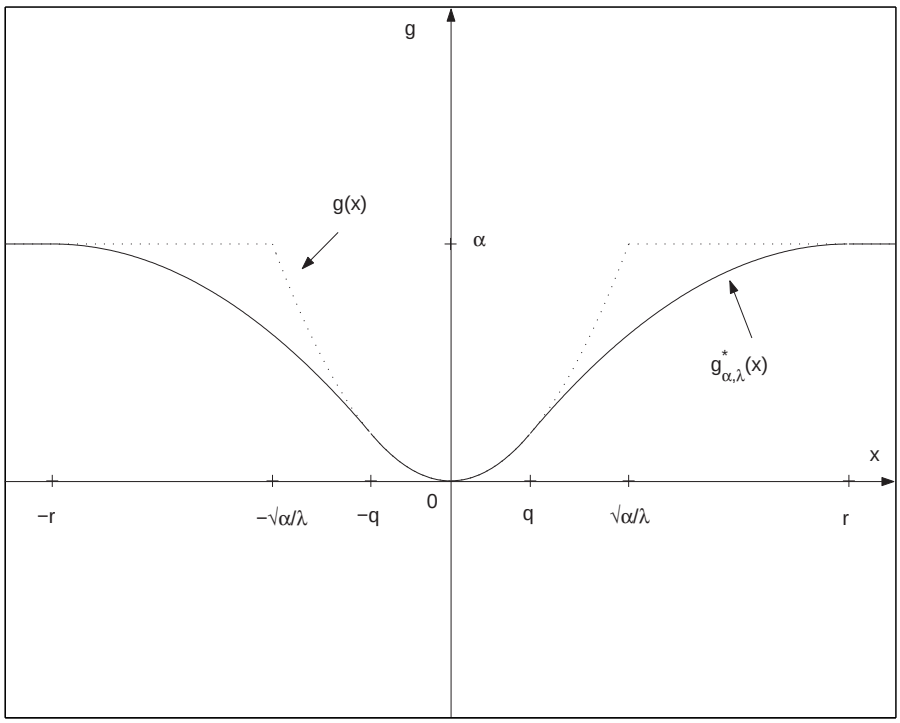

Figura 6.1: Grafico della $g^*_{\alpha,\lambda}$ del funzionale convesso che approssima la g del funzionale iniziale.

6.1.1 Caso della weak membrane

Nel caso D dimensionale del modello discretizzato di energia (A.3) il funzionale ha un'espressione equivalente

$$E(u) = D + \sum_k g(\sum_l Q_{k,l} u_l) \tag{6.5}$$

con $Q = (Q_{k,l})$ una opportuna matrice detta *circolante*, perchè presenta righe in cui si ripete un vettore in posizioni diverse.

Osservazione 6.1 *Nel caso 1D la matrice Q, definita in (A.5), ha la coppia $(-1, 1)$ che si ripete in ogni riga (tranne che nella prima) e vale*

l'uguaglianza

$$\forall k = 2, \ldots, N \ \sum_{l=1}^{N} Q_{k,l} u_l = u_k - u_{k-1}$$

che prova l'equivalenza fra la (6.5) e la (6.1). La (6.1) considerando una discretizzazione in $\{x_1, \ldots, x_N\}$ *ha la forma più precisa*

$$E = \sum_{i=1}^{N} (u_i - d_i)^2 + \sum_{k=2}^{N} g(u_k - u_{k-1}) = D + \sum_{k=2}^{N} g(\sum_{l=1}^{N} Q_{k,l} u_l),$$

perché non ha senso per $k = 1$ *la sottrazione* $u_k - u_{k-1}$ *visto che* u_0 *non è definibile.*

Nell'appendice A.2.1 viene determinata un'espressione del tipo (6.5) nel caso 2D della weak membrane (6.2).

Ora, se si vuole determinare il funzionale convesso nel modello della weak membrane si sfrutta una condizione simile a quella per la weak string (6.3), dove ora Q è una matrice tale che $Q^T Q = P^T P + R^T R$, con P e R espresse nell'appendice.

Allora varrà la seguente proposizione per il modello della weak membrane e nel caso generale D dimensionale

Proposizione 6.1.1 *Se la* $g(x, y) = g_1(x) + g_1(y)$ *presente nel modello (6.2) soddisfa la condizione:*

$$\exists c^* > 0 \text{ tale che } \mathrm{g_1''(x) + g_1''(y) \geq -c^* \text{ e } v_{max} \leq 2/c^*},$$

dove $v_{max} = 8$ *è il massimo autovalore di* $P^T P + R^T R$*, allora il funzionale dell'energia (6.2) con tale* $g(x, y)$ *è convesso. Mentre nel caso generale* D *dimensionale (6.5) la condizione sufficiente alla convessità è*

$$\exists c^* > 0 \text{ tale che } \mathrm{g''(x) \geq -c^* \text{ e } v_{max} \leq 2/c^*},$$

con v_{max} *il massimo autovalore di* $Q^T Q$*.*

Grazie a questa proposizione si può provare che il seguente funzionale è convesso

$$g^*_{\alpha,\lambda}(x,y) = \begin{cases} \lambda^2|(x,y)| & se \quad |(x,y)| < q \\ \alpha - c^*\dfrac{(|(x,y)| - r)^2}{2} & se \quad q \leq |(x,y)| < r \\ \alpha & se \quad |(x,y)| \geq r \end{cases} .$$

dove come prima

$$r^2 = \alpha\left(\frac{2}{c^*} + \frac{1}{\lambda^2}\right), \quad e \quad q = \frac{\alpha}{\lambda^2 r} .$$

Esso è quello di partenza nell'algoritmo GNC secondo B&Z applicato al modello weak membrane (6.2).

Osservazione 6.2 *Nella parte successiva del capitolo viene analizzato il modello della weak string ma risultati analoghi valgono e vengono estesi anche sulla weak membrane.*

6.1.2 Sufficienza del primo passo

A volte la determinazione del funzionale convesso $E^*(u)$ può bastare per determinare il minimo di $E(u)$ senza dover applicare il secondo passo dell'algoritmo GNC. Ciò avviene quando $E^*(u)$ è una buona approssimazione di $E(u)$ tale che il suo minimo u^* è proprio il minimo di $E(u)$.

Si può osservare che quanto più λ diminuisce allora tanto più $g^*(x)$ sarà una migliore approssimazione di $g(x)$ (si confronti la Figura 6.2 con la 6.1) e quindi E^* sarà una migliore approssimazione di E.

In effetti si ha che q e r al diminuire di λ si avvicinano e tendono a $\frac{\sqrt{\alpha}}{\lambda}$ per $\lambda \to 0$ facendo così tendere $g^*_{\alpha,\lambda}$ a g e quindi u^* al minimo di $E(u)$. Siccome però il parametro λ è fissato a priori e non può essere modificato, si sono determinate delle condizioni più operative

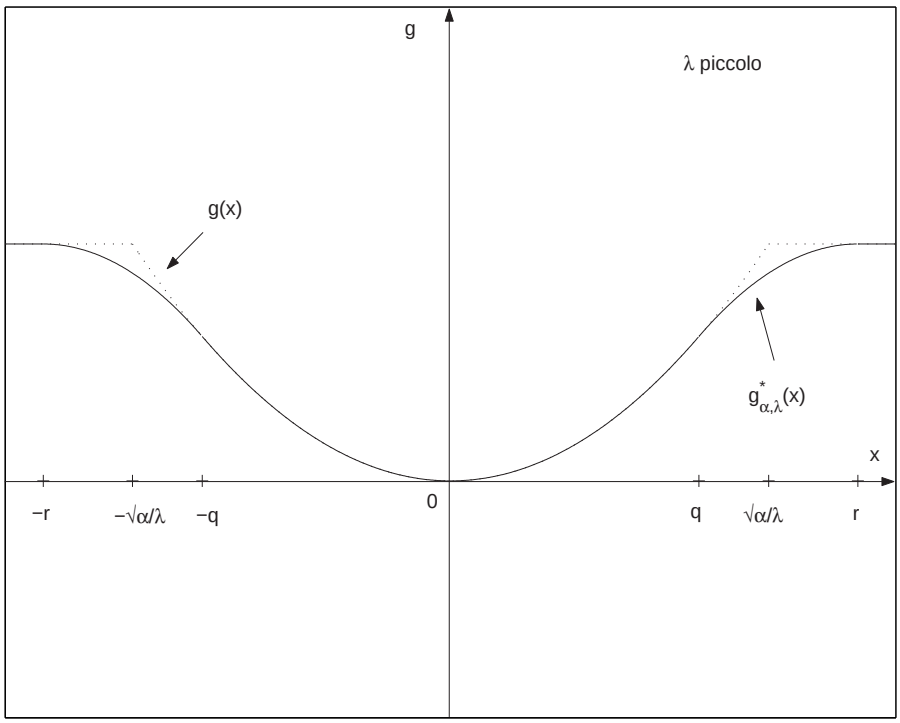

Figura 6.2: Grafico della $g^*_{\alpha,\lambda}$ del funzionale convesso nel caso λ piccolo. Si noti che esso approssima meglio la g del funzionale iniziale rispetto a quella rappresentata in figura 6.1 che ha un parametro λ più alto.

che garantiscono la sufficienza del primo passo e quindi un notevole risparmio dei costi computazionali.

Una condizione sufficiente affinche u^* è anche il minimo di $E(u)$ è

$$E^*(u^*) = E(u^*). \tag{6.6}$$

Infatti, ricordando che $\forall x\ g^*(x) \leq g(x)$, si ottiene che

$$\forall u = u(t)\ E^*(u) \leq E(u)$$

e quindi se è vera la (6.6) si ha $\forall u = u(t)$

$$E(u^*) = E^*(u^*) \leq E^*(u) \leq E(u).$$

Si è provato in [2] che se una stringa presenta un solo salto centrale di altezza h che non appartiene all'intervallo $[h_-, h_+] = [h_0/2\lambda, h_0\sqrt{2\lambda}]$ allora la condizione (6.6) è soddisfatta. Si noti che anche in questa situazione per λ piccolo (vicino a 1) l'intervallo diventa molto piccolo e quindi solo il primo passo può determinare la ricostruzione ottima.

Osservazione 6.3 *La condizione (6.6) è soddisfatta solo se*

$$\forall i \; g^*_{\alpha,\lambda}(\Delta^*_i) = g(\Delta^*_i),$$

dove $\Delta^*_i = u^*_i - u^*_{i-1}$, *e quindi se ogni* Δ^*_i *verifica* $|\Delta^*_i| \leq q \vee |\Delta^*_i| \geq r$. *La prima disuguaglianza ci dice che nel funzionale* E^* *la* $g^*(\Delta^*_i)$, *per la (6.4), è del tipo* $\lambda^2\Delta^*_i$. *Quindi nel modello dell'energia* E^* *nell'intervallo* $[x_{i-1}, x_i]$ *si forza la continuità nello stesso spirito che è proprio dei line process (3.2), dove per* $l_i = 1$, *si forza la continuità eliminando la penalizzazione* α. *Mentre nel caso* $|\Delta^*_i| \geq r$ *il pezzo nel termine* S *corrispondente è* α *che indica nella ricostruzione la presenza di una discontinuità in* $[x_{i-1}, x_i]$, *cioè nell'"ipotetico" modello line process di* E^* *vale* $l_i = 0$.

Questo modello di line process per $E^*(u)$ *non esiste (è "ipotetico") per una* u *qualsiasi che non soddisfa la (6.6). Infatti, in tal caso, esiste qualche* i *tale che* $q < |\Delta_i| < r$, *ed allora le ipotetiche* l_i *avrebbero un valore incerto. L'intervallo* $[q, r]$, *si dice per l'appunto una zona di ambiguità per le* Δ_i.

6.2 Secondo passo dell'algoritmo GNC

L'osservazione 6.3 del paragrafo precedente mette in luce che, nel caso in cui la (6.6) non è soddisfatta (cioé per λ sufficientemente grande), è necessario introdurre una sequenza di funzionali $E^{(p)}$ con una zona di ambiguità sempre più ristretta.

Tale sequenza verrà poi bloccata quando la ricostruzione minima $u^{(p)}$ di $E^{(p)}$ produce le

$$\Delta_i^{(p)} = u_i^{(p+1)} - u_i^{(p)}$$

che cadono tutte fuori da tale zona, perchè in tal caso, troveremo $E^{(p)}(u^{(p)}) = E^*(u^{(p)})$ che ci dice che $u^{(p)}$ è proprio il minimo di $E^{(p)}$, visto che le $g^{(p)}$ che costruiremo si troveranno sempre al di sotto di g.

La successione dei funzionali dell'energia $\{E^{(p)}\}_{p\in]0,1]}$ secondo gli appunti di B&Z [2] è così definita $\forall p \in]0,1]$

$$E^{(p)}(u,d) = E^{(p)}(u) = D + \sum_i g^{(p)}(u_i - u_{i-1}).$$

con

$$g^{(p)}(x) = g^{(p)}_{\alpha,\lambda}(x) = \begin{cases} \lambda^2 x^2 & se \quad |x| < q^{(p)} \\ \alpha - c^{(p)}\dfrac{(|x| - r^{(p)})^2}{2} & se \quad q^{(p)} \le |x| < r^{(p)} \\ \alpha & se \quad |x| \ge r^{(p)} \end{cases} . \tag{6.7}$$

dove

$$c^{(p)} = \frac{c^*}{p}, \quad (r^{(p)})^2 = \alpha\left(\frac{2}{c^{(p)}} + \frac{1}{\lambda^2}\right), \quad e \quad q^{(p)} = \frac{\alpha}{\lambda^2 r^{(p)}} .$$

Quindi otteniamo per la (6.4) $g^*_{\alpha,\lambda} = g^{(1)}_{\alpha,\lambda}$ e $E^{(1)} = E^*$. Si osserva che la successione $\{E^{(p)}\}$ tende per $p \to 0$ a $E(u)$, il modello iniziale (6.1) dell'energia. Infatti siccome $c^{(p)} \to +\infty$ si ha che $r^{(p)}, q^{(p)} \to \frac{\sqrt{\alpha}}{\lambda}$ e quindi $g^{(p)}(x) \to g(x)$. Allora si pone $g^{(0)} \equiv g$ e $E^{(0)} \equiv E$. Allora effettivamente le $E^{(p)}$ approssimano sempre meglio il funzionale iniziale partendo da quello convesso E^* a $p = 1$, ma ciononostante resta da provare che i minimi di $E^{(p)}$ tendono al minimo $\widetilde{u}$ del funzionale iniziale E. Questi minimi vengono determinati in ogni passo p applicando un metodo di discesa che parti dal minimo del passo precedente. Si parla di passo precedente perché nella pratica le p passeranno da 1 a 0 in modo discreto con ad esempio un salto fra due p consecutive $p_k - p_{k-1}$ costante.

6.3 Correttezza dell'algoritmo GNC

Ora verifichiamo se l'algoritmo GNC impostato funziona correttemente, cioé se determina il minimo di $E(u)$ in un numero finito di passi. Iniziamo con l'esaminare il caso delle discontinuità isolate.

Supponiamo di avere dei segnali $d(x)$ discretizzati in $x_{-L}, \ldots, x_L$ cioé $d_{-L}, \ldots, d_L$ e che è presente una sola discontinuità in 0, $h = |d_1 - d_{-1}|$. Supponiamo inoltre che la ricostruzione convessa u^* presenti il solo $\Delta_0^* = u_1^* - u_{-1}^*$ nella zona ambigua $]q, r[$, mentre gli altri Δ_i^* si trovino già nella zona di ricostruzione continua, cioé $|\Delta_i^*| < q$.

Come sappiamo questo salto h nei dati, può portare o ad una ricostruzione finale $\widetilde{u}$ continua oppure ad un salto in 0 per la caratteristica del modello di weak continuity di E a seconda che $h < h_0$ o meno; ora invece vediamo come, la scelta del modello da utilizzare si sviluppa, alla variazione del parametro p, nel secondo passo dell'algoritmo. Se facciamo progredire l'algoritmo GNC al diminuire di p, $\Delta_0^{(p)}$ varia con continuità, così come i limiti della zona di ambiguità $[q^{(p)}, r^{(p)}]$ che tendono ad avvicinarsi e si incontrano per $p = 0$. Allora, siccome $\Delta_0^{(p)}$ parte da $\Delta_0^* = \Delta_0^{(1)}$, che si trova in $]q, r[=]q^{(1)}, r^{(1)}[$, essa resta sempre in $]q^{(p)}, r^{(p)}[$ fino a che non interseca per un certo $\overline{p}$ la $q^{(\overline{p})}$ o $r^{(\overline{p})}$ (Figura 6.3).

Nel primo caso la ricostruzione $u^{(\overline{p})}$ è continua in [-1,1], nell' altro viene forzata nel modello la discontinuità. Allora avremo che tutte le $\Delta_i^{(\overline{p})}$ si trovano fuori dalla zona di ambiguità, e quindi risulta

$$E^{(\overline{p})}(u^{(p)}) = E(u^{(\overline{p})}),$$

cioé $u^{(\overline{p})}$ è la ricostruzione ottima $\widetilde{u}$ di E.

Questa è solo una dimostrazione grafica della correttezza del secondo passo, mentre la successiva è più analitica.

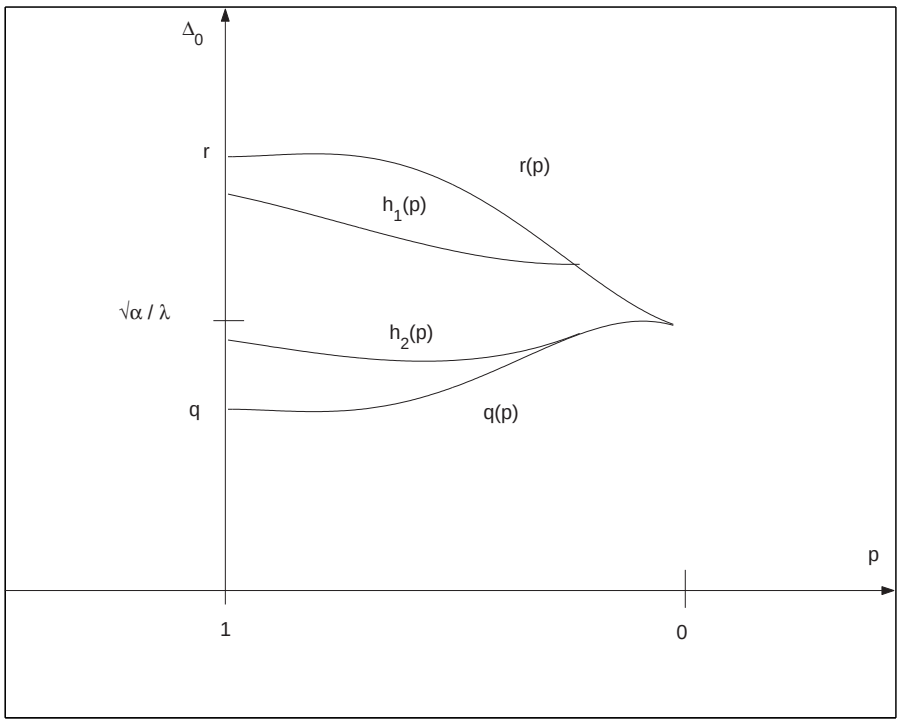

Figura 6.3: Se partiamo da $\Delta_0 = h_1(0)$ vicino a r la $h_1(p) = \Delta_0^{(p)}$ tende ad incontrare $r(p)$. Allora in tal caso la ricostruzione finale inserisce una discontinuità in 0. Mentre nel caso di $\Delta_0^{(p)} = h_2(p)$ il modello E forza la continuità in 0.

Supponiamo che i segnali a disposizione siano del tipo

$$d_i = \begin{cases} h/2 & se \quad i > 0 \\ -h/2 & se \quad i < 0 \end{cases} .$$

Allora con le stesse ipotesi precedenti si ha che, ad un generico passo p, il funzionale $E^{(p)}$, se calcolato nel suo minimo $u^{(p)}$, assume un valore dipendente dalla sola $\Delta_0^{(p)} > 0$ che si indica con $E^{(p)}(\Delta_0^{(p)})$ e che vale

$$E^{(p)}(\Delta_0^{(p)}) = \frac{1}{2}\mathcal{F}(h - \Delta_0^{(p)})^2 + g^{(p)}(\Delta_0^{(p)}), \tag{6.8}$$

con $\mathcal{F}$ una costante positiva. Tale $E^{(p)}$ è convessa (in Δ_0) se e solo se $c^{(p)} \leq \mathcal{F}$. In particolare, se $c^{(p)} = \mathcal{F}$ avremo che $E^{(p)}(\Delta_0^{(p)})$ ha la derivata seconda nulla

$$\frac{d^2 E^{(p)}}{d\Delta_0^2} = 0$$

in tutto $[q^{(p)}, r^{(p)}]$, ossia in tale intervallo $E^{(p)}$ è lineare; mentre se $c^{(p)} > \mathcal{F}$ la concavità è verso il basso (Figura 6.4).

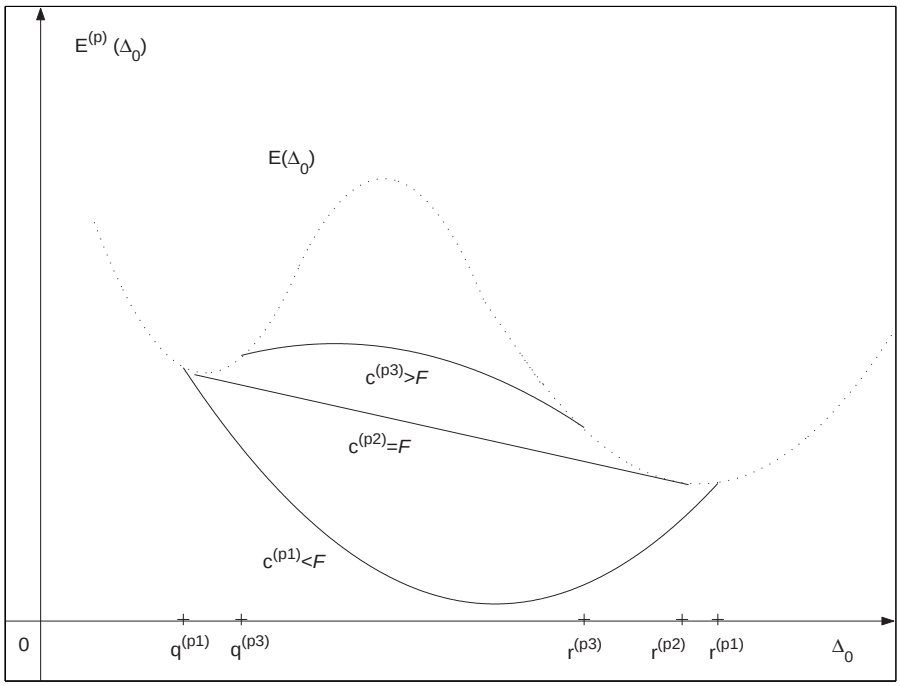

Figura 6.4: Siccome $c^{(p1)} < \mathcal{F}$ si ha $E^{(p1)}$ è convessa. Ció conduce ad un minimo in $]q^{(p1)}, r^{(p1)}[$ che non è il minimo di E. Mentre se prendo $p_2 < p_1$ tale che $p_2 = c^*/\mathcal{F}$ allora $E^{(p2)}$ è lineare in $]q^{(p1)}, r^{(p1)}[$, il minimo di $E^{(p)}$ è allora quello di E. Se invece $p_3 > \mathcal{F}$ il metodo di discesa può condurre a un minimo solo locale di E.

Chiaramente nel caso convesso (con p sufficientemente grande, tale

che $p \geq c^*/\mathcal{F}$), è applicabile un metodo di discesa per determinare il minimo di $E^{(p)}$. Questo minimo sarebbe anche quello di E (e quindi avremmo terminato l'algoritmo), se si trovasse fuori dall'intervallo $]q^{(p)}, r^{(p)}[$. Ció succede sicuramente nel caso lineare, in cui il minimo di $E^{(p)}$ in $[q^{(p)}, r^{(p)}]$ si ottiene per $\Delta_0^{(p)} = q^{(p)}$ o $\Delta_0^{(p)} = r^{(p)}$ e quindi il minimo globale di $E^{(p)}$ si trova fuori $]q^{(p)}, r^{(p)}[$. Allora la ricostruzione $u^{(p)}$, relativa alla Δ_0 che minimizza $u^{(p)}$ ovunque, sarà tale che $E^{(p)}(u^{(p)}) = E(u^{(p)})$, cioé $u^{(p)}$ minimizza anche l'energia iniziale E . Però questo caso lineare, in cui $c^{(p)} = \mathcal{F}$, descrive una situazione instabile per il funzionale $E^{(p)}$ e quindi si preferisce utilizzare, nella pratica, una configurazione di $E^{(p)}$ con una $c^{(p)} > \mathcal{F}$ che porta sempre al minimo di E garantendo in più la stabilità. Tale parametro $c^{(p)}$ non deve essere troppo alto, altrimenti il metodo di discesa su $E^{(p)}$ potrebbe condurre ad un minimo locale di E.

Si può provare che per $L >> \lambda$ risulta $\mathcal{F} \leq \lambda$, e che per $c^{(p)}$ poco più grande di λ, cioé per p poco più piccolo di c^*/λ, si può cadere in un altro minimo locale (Figura 6.4), mentre per $c^(p) = \lambda$, si raggiunge il minimo assoluto di E senza problemi di stabilità. In conclusione allora, viene scelta come p finale dell'algoritmo GNC la $p = \dfrac{c^*}{\lambda}$ che corrisponde a $c^{(p)} = \lambda$.

Osservazione 6.4 *Nel paragrafo 5.1 è stato provato, attraverso l'analisi variazionale del funzionale dell'energia della weak string E, che la ricostruzione che minimizza l'energia prevede la discontinuità isolata (in 0) se il suo salto h supera il limite h_0. Ora, con l'applicazione di questo algoritmo GNC, si ha un risultato simile infatti, si prova che il limite di sensitività risulta essere pari a $h_0 + O(1/\lambda^2)$*

Dall'analisi della figura 6.4 si osserva che al diminuire di p non è detto che il minimo di $E^{(p)}$ si comporta monotonicamente decrescendo fino al

minimo di E, infatti per $p = p_1$ ho un minimo di $E^{(p)}$, minore di quello assoluto di E.

6.3.1 Discontinuità interagenti

Supponiamo ora che nei dati siano presenti due discontinuità: di salto h_1 in 0 e h_2 in $x = a$. Allora vogliamo vedere se l'algoritmo GNC permette anche in questo caso la determinazione del minimo di E, rispettando i criteri variazionali stabiliti nel paragrafo 5.1.1.
Supponendo di partire da una energia convessa E^* che abbia le $\Delta_i^* < q$ per $i \neq 0, a$; si ottiene che la generica $E^{(p)}(u^{(p)})$ (calcolata nel suo minimo $u^{(p)}$) dipende dai salti $\Delta_0^{(p)}$ e $\Delta_a^{(p)}$:

$$\begin{aligned} E^{(p)}(\Delta_0^{(p)}, \Delta_a^{(p)}) \quad = \quad & \tfrac{\lambda}{2}(\Delta_0^{(p)} - h_1)^2 + \tfrac{\lambda}{2}(\Delta_a^{(p)} - h_2)^2 \\ & -\lambda \exp(-\tfrac{a}{\lambda})(\Delta_0^{(p)} - h_1)(\Delta_a^{(p)} - h_2) \\ & +g^{(p)}(\Delta_0^{(p)}) + g^{(p)}(\Delta_a^{(p)}). \end{aligned}$$

Stiamo supponendo che $a << \lambda$ perchè altrimenti otterremmo il caso precedente delle discontinuità isolate. Nel caso del cappello a cilindro con $h_1 = h_2$ si ha che $\Delta_0^{(p)} = \Delta_a^{(p)}$ e quindi

$$E^{(p)} = E^{(p)}(\Delta_0^{(p)}) = 2\left(\frac{a}{2}(\Delta_0^{(p)} - h)^2 + g^{(p)}(\Delta_0^{(p)})\right); \qquad (6.9)$$

dove si è sfruttata l'approssimazione valida per $a << \lambda$: $1 - exp(-a/\lambda) \simeq a/\lambda$. Il modello (6.9) è simile al (6.8) visto per le discontinuità isolate, perciò si ha che:

- se $c^{(p)} < a$ la $E^{(p)}$ è convessa in $[q^{(p)}, r^{(p)}]$, per cui si potrebbe determinare un minimo per $\Delta_0^{(p)} \in]q^{(p)}, r^{(p)}[$ che non sarebbe minimo di E.

- se $c^{(p)} = a$ si ha l'instabilità, la $E^{(p)}$ è lineare in $[q^{(p)}, r^{(p)}]$ e si determina esattamente il minimo di E.

- se $c^{(p)} = \lambda >> a$ la soluzione minima è stabile e coincide con quella di E.

Quindi in questa situazione la p finale da prendere é $p = c^*/\lambda$.

Nel caso $h_1 < h_2$ si determinano due stati instabili

$$c^{(p1)} = \lambda\left(1 - \exp\left(-\frac{a}{\lambda}\right)\right),\ c^{(p2)} = a$$

ognuno associato ad una discontinuità. Allora scelta $p = \min\{p_1 p_2\}$, avremo che la coppia $(\Delta_0^{(p)}, \Delta_a^{(p)})$ che minimizza l'energia $E^{(p)}$ si troverà fuori dalla zona di ambiguità e quindi il suo minimo coinciderà con quello dell'energia iniziale E. La configurazione della ricostruzione ottenuta (nessuna,una o due discontinuità) dipenderà da quale delle tre energia viste alla fine del paragrafo paragrafo 5.1.1 è minore.

6.3.2 Rumore

Nel capitolo 5 abbiamo visto che, per garantire l'immunità dal rumore, deve essere garantita nel modello che $\alpha > 2\sigma^2$ dove σ è la deviazione standard della sorgente di rumore. Con l'applicazione dell' algoritmo GNC risulta che l'immunità si ottiene se è soddisfatta una relazione meno restringente:

$$\alpha > \frac{2\sigma^2}{\lambda}.$$

6.4 Individuazione dell'algoritmo

La scelta dell'algoritmo di discesa di $E^{(p)}$ deve chiaramente basarsi sul tipo di funzioni da minimizzare Nel seguito si suppone di operare su funzionali $E^{(p)}(u_0, \ldots, u_N)$ tali che le funzioni

$$u_l \longrightarrow F_l^{(p)}(u_l) \equiv E^{(p)}(\overline{u_0}, \ldots, u_l, \ldots, \overline{u_N})$$

(dove le $\overline{u_i}$ sono fissate) sono delle quadratiche nell'intorno del minimo. Consideriamo allora un algoritmo di discesa del gradiente che utilizza uno schema di sovrarilassamento non lineare di questo tipo: $\forall\ l, n$

$$u_l^{(n+1)} = u_l^{(n)} - \frac{\omega}{T_l}\frac{\partial E^{(p)}}{\partial u_l}(u^{(n)}) = u_l^{(n)} - \frac{\omega}{T_l}\frac{dF_l^{(p)}}{du_l}(u_l^{(n)}). \tag{6.10}$$

Esso permette di raggiungere il minimo $u^{(p)} = (u_0^{(p)}, \ldots, u_N^{(p)})$ della funzione $E^{(p)}$. Infatti, scelto dal ciclo esterno il punto l, si opererà al variare di n per costruire la successione $u_l^{(n)}$ che si avvicinerà, come proveremo di seguito, al minimo di $E^{(p)}$ calcolato in l, cioé a $u_l^{(p)}$. Basterà allora provare la convergenza al minimo di $F_l^{(p)}$ (in cui supponiamo che per $i \neq l$ $\overline{u_i} = u_i^{(p)}$). Il parametro ω è quello di rilassamento ($0 < \omega < 2$, per avere la convergenza), mentre $T_l > 0$ è un limite superiore della derivata seconda di $E^{(p)}$ cioé vale

$$\frac{\partial^2 E^{(p)}}{\partial u_l^2} \leq T_l. \tag{6.11}$$

Per provare la convergenza, o meglio la tendenza al minimo di $F_l^{(p)}$, supponiamo di partire da un punto $u_l^{(n)}$ che si trovi nella concavità a cui appartiene il minimo di $F_l^{(p)}$, con $\frac{dF_l^{(p)}}{du_l}(u_l^{(n)}) < 0$ e $\frac{d^2F_l^{(p)}}{du_l^2} > 0$. Allora per l'iterazione (6.10) si ottiene $u_l^{(n)} < u_l^{(n+1)}$.

Nelle prossime righe verificheremo se la $u_l^{(n+1)}$, ottenuta dal metodo di iterazione, si avvicina al minimo di $F_l^{(p)}$, ossia se $\frac{dF_l^{(p)}}{du_l}(u_l^{(n+1)})$ è più vicina a zero rispetto alla derivata in $u_l^{(n)}$.

Partiamo dal caso $\omega \leq 1$. Per il teorema del valor medio sappiamo che $\forall\ u_l \in [u_l^{(n)}, u_l^{(n+1)}]\ \exists \xi \in [u_l^{(n)}, u_l]$ tale che

$$\frac{dF_l^{(p)}}{du_l}(u_l) - \frac{dF_l^{(p)}}{du_l}(u_l^{(n)}) = \frac{d^2F_l^{(p)}}{du_l^2}(\xi)(u_l - u_l^{(n)}).$$

Siccome $\frac{d^2F_l^{(p)}}{du_l^2} = \frac{\partial^2 E^{(p)}}{\partial u_l^2}$ e $u_l \leq u_l^{(n+1)}$, avremo per la (6.11) che

$$\frac{dF_l^{(p)}}{du_l}(u_l) - \frac{dF_l^{(p)}}{du_l}(u_l^{(n)}) \leq T_l(u_l^{(n+1)} - u_l^{(n)}).$$

Per l'iterazione (6.10) si ottiene

$$\frac{dF_l^{(p)}}{du_l}(u_l) - \frac{dF_l^{(p)}}{du_l}(u_l^{(n)}) \leq T_l\left(-\frac{\omega}{T_l}\frac{dF_l^{(p)}}{du_l}(u_l^{(n)})\right) = -\omega\frac{dF_l^{(p)}}{du_l}(u_l^{(n)}).$$

Se isoliamo al primo membro la derivata in u_l vale la relazione

$$\frac{dF_l^{(p)}}{du_l}(u_l) \leq (1-\omega)\frac{dF_l^{(p)}}{du_l}(u_l^{(n)}). \tag{6.12}$$

Da cui siccome $0 \leq 1-\omega < 1$ si ha che $(1-\omega)\frac{dF_l^{(p)}}{du_l}(u_l^{(n)}) \leq 0$ e quindi $\frac{dF_l^{(p)}}{du_l}(u_l)$ resta negativo ma più vicino a zero rispetto a $u_l^{(n)}$, infatti per la crescenza della derivata nella quadratica avremo che

$$0 \geq (1-\omega)\frac{dF_l^{(p)}}{du_l}(u_l^{(n)}) \geq \frac{dF_l^{(p)}}{du_l}(u_l) \geq \frac{dF_l^{(p)}}{du_l}(u_l^{(n)}).$$

Ad esempio se $\frac{dF_l^{(p)}}{du_l}(u_l^{(n)}) = -2$ e $\omega = 0.5$ la derivata in un punto qualsiasi di $[u_l^{(n)}, u_l^{(n+1)}]$ può variare in $[-2,-1]$. Per la monotonia delle derivate avremo che la $u_l = u_l^{(n+1)}$ si avvicina maggiormente al minimo rispetto a tutte le $u_l \in [u_l^{(n)}, u_l^{(n+1)}]$, avendo quindi una $F_l^{(p)}$ minore di quella in $u_l^{(n)}$.

Se invece $\omega > 1$ allora $1-\omega < 0$ e quindi per la (6.12) e la crescenza della derivate si ha $\forall u_l \in [u_l^{(n)}, u_l^{(n+1)}]$

$$\frac{dF_l^{(p)}}{du_l}(u_l^{(n)}) \leq \frac{dF_l^{(p)}}{du_l}(u_l) \leq \underbrace{(1-\omega)\frac{dF_l^{(p)}}{du_l}(u_l^{(n)})}_{>0}.$$

Quindi in questo caso nell'intervallo può cambiare il segno della derivata, cioè $u_l^{(n+1)}$ può aver superato il punto di minimo di $F_l^{(p)}$. Però solo nel caso $\omega \geq 2$ in cui si ha $1 - \omega \leq -1$ l'allontanamento dal punto di minimo è tale da portare in $u_l^{(n+1)}$ un incremento dell'energia rispetto a quella di $u_l^{(n)}$.

6.4.1 Scelta del parametro di rilassamento ω

Si prova che la diminuzione maggiore dell'energia $F_l^{(p)}$ si ottiene per $\omega = 1$, ma per l'uso combinato dell'algoritmo per ogni l la convergenza più veloce verso il minimo globale di $E^{(p)}$, si ottiene per un $1 < \omega < 2$, si veda [2]. Precisamente l'omega ottimale è nei due casi di continuità debole:

- weak string
$$\omega = \frac{2}{(1 + 1/\lambda)}$$
- weak continuity
$$\omega = \frac{2}{\left(1 + \dfrac{1}{\sqrt{2}\lambda}\right)}. \tag{6.13}$$

6.4.2 Implementazione dell'algoritmo

Per poter implementare l'algoritmo retto dall'iterazione (6.10) dobbiamo rappresentare meglio la derivata parziale $\dfrac{\partial E^{(p)}}{\partial u_l}$ e determinare un maggiorante della derivata seconda. Consideriamo il modello della $E^{(p)}$, espresso nella forma generale del tipo (6.5), ovvero con la particolare norma robusta $g_{\alpha,\lambda}^{(p)}$ definita in (6.7)

$$E^{(p)}(u) = \sum_i (u_i - d_i)^2 + \sum_k g_{\alpha,\lambda}^{(p)} \left(\sum_s Q_{k,s} u_s \right), \tag{6.14}$$

grazie al quale potremo dedurre l'algoritmo da implementare nel caso della weak string e della weak membrane.

Dal modello (6.14) determiniamo la derivata parziale

$$\frac{\partial E^{(p)}}{\partial u_l}(u) = 2(u_l - d_l) + \sum_k {g^{(p)}_{\alpha,\lambda}}' \left(\sum_s Q_{k,s} u_s \right) Q_{k,l} \tag{6.15}$$

con

$${g^{(p)}_{\alpha,\lambda}}'(x) = \begin{cases} 2\lambda^2 x & \text{se } |\mathrm{x}| < \mathrm{q^{(p)}} \\ -c^{(p)}(|x| - r^{(p)}) \cdot sign(x) & \text{se } \mathrm{q^{(p)}} \le |\mathrm{x}| < \mathrm{r^{(p)}} \\ 0 & \text{se } |\mathrm{x}| \ge \mathrm{r^{(p)}} \end{cases}, \tag{6.16}$$

per la definizione (6.7). Mentre la derivata seconda di $E^{(p)}$ è

$$\frac{\partial^2 E^{(p)}}{\partial u_l^2}(u) = 2 + \sum_k {g^{(p)}_{\alpha,\lambda}}'' \left(\sum_s Q_{k,s} u_s \right) Q^2_{k,l}$$

dove, per la (6.16)

$${g^{(p)}_{\alpha,\lambda}}''(x) = \begin{cases} 2\lambda^2 & \text{se } |\mathrm{x}| < \mathrm{q^{(p)}} \\ -c^{(p)} \cdot (sign(x))^2 = -c^{(p)} & \text{se } \mathrm{q^{(p)}} < |\mathrm{x}| < \mathrm{r^{(p)}} \\ 0 & \text{se } |\mathrm{x}| \ge \mathrm{r^{(p)}} \end{cases}.$$

Allora $\forall x \ {g^{(p)}_{\alpha,\lambda}}''(x) \le 2\lambda^2$ e quindi otteniamo che un maggiorante T_l di $\dfrac{\partial^2 E^{(p)}}{\partial u_l^2}(u)$ è

$$T_l = 2 + 2\lambda^2 \sum_k Q^2_{k,l}. \tag{6.17}$$

A questo punto possiamo esprimere l'iterata n-sima (6.10) nel punto l nel modello weak string e in quello della weak membrane che verrà analizzato nel successivo capitolo.

Nel caso del modello weak string supponiamo di discretizzare la stringa in N punti $\{x_1, \ldots, x_N\}$ allora la forma dell'energia $E^{(p)}$ è per la (6.14), vedi l'Osservazione 6.1

$$E^{(p)} = \sum_{i=1}^{N} (u_i - d_i)^2 + \sum_{k=2}^{N} g^{(p)}_{\alpha,\lambda} (\sum_{s=1}^{N} Q_{k,s} u_s)$$

da cui otteniamo che, per la (6.15), l'espressione dell'iterata (6.10) diventa $\forall l = 1, \ldots, N$

$$u_l^{(n+1)} = u_l^{(n)} - \frac{\omega}{T_l}\left(2(u_l^{(n)} - d_l) + \sum_{k=2}^{N} g_{\alpha,\lambda}^{(p)\prime}\left(\sum_{s=1}^{N} Q_{k,s} u_s^{(n)}\right) Q_{k,l}\right) \quad (6.18)$$

Questa iterata n-sima sul calcolatore viene implementata dentro i tre cicli:

$$\begin{array}{l} for\ p = 1, p_2, p_3, ...p_{min} = 1/(2\lambda) \\ \quad for\ l = 1, ..., N \\ \quad\quad while\ |u_l^{(n+1)} - u_l^{(n} | >= precisione \\ \quad\quad\quad n \equiv n + 1 \\ \quad\quad\quad .....iterata\ nsima..... \\ \quad\quad end; \\ \quad end; \\ end; \end{array} \quad (6.19)$$

Nel ciclo esterno si decrementa p in modo discreto, mediante una sequenza $\{p_k\}$, da 1 a fino a $p_{min} = \dfrac{c^*}{\lambda} = \dfrac{1}{2\lambda}$, secondo quanto detto nel paragrafo 6.3. Nell'altro ciclo si incrementa la coordinata l del dominio, iniziando da $l = 1$ e fino all'altro estremo del dominio $l = N$. Infine nel terzo ciclo si applica il metodo di discesa sulla funzione componente l-sima di $E^{(p)}$ ($u_l \longrightarrow E^{(p)}$), con l'iterata n-sima uguale alla (6.18), per determinare una approssimazione della coordinata l del minimo di $E^{(p)}$ (che abbiamo indicato per abuso di notazione con $u_l^{(p)}$) entro una *precisione* fissata a priori. Il punto di partenza di ogni ciclo interno per una data coordinata l, è $u_l^{(n=0)}$ pari al minimo di $E^{(p_{k-1})}$ ottenuto al passo precedente dove $p = p_{k-1}$.

L'iterata può essere semplificata nei tre casi $l = 1$, $1 < l < N$, $l = N$. Infatti, ad esempio, il coefficiente T_l definito in (6.17) si comporta in maniera diversa

- $T_l = 2 + 2\lambda^2$ se $l = 1$ e $l = N$

- $T_l = 2 + 4\lambda^2 \quad \text{se } 1 < l < N.$

Inoltre anche la sommatoria ha un espressione diversa nei tre casi. Iniziamo a considerare il caso $l = 1$. Per la definizione della matrice $Q \in \mathbb{R}^{N\times N}$ (data in (A.5)), si ha che (al variare di k in $\{2, \ldots, N\}$) solo per $k = 2$ risulta $Q_{k,1} \neq 0$ per cui

$$\sum_{k=2}^{N} {g_{\alpha,\lambda}^{(p)}}' \left(\sum_{s=1}^{N} Q_{k,s} u_s^{(n)} \right) Q_{k,1} = -{g_{\alpha,\lambda}^{(p)}}' \left(\sum_{s=1}^{N} Q_{2,s} u_s^{(n)} \right)$$

e siccome gli unici $Q_{2,s} \neq 0$ sono $Q_{2,1} = -1$ e $Q_{2,2} = 1$ otteniamo l'iterata

$$u_1^{(n+1)} = u_1^{(n)} - \frac{\omega}{2 + 2\lambda^2} \left(2(u_1^{(n)} - d_1) - {g_{\alpha,\lambda}^{(p)}}'(u_2^{(n)} - u_1^{(n)}) \right).$$

che diventa

$$u_1^{(n+1)} = u_1^{(n)} - \frac{\omega}{2 + 2\lambda^2} \left(2(u_1^{(n)} - d_1) + {g_{\alpha,\lambda}^{(p)}}'(u_1^{(n)} - u_2^{(n)}) \right).$$

giacché la ${g_{\alpha,\lambda}^{(p)}}'$ è dispari.

Osservazione operativa 6.5 *Visto che stiamo lavorando per $l = 1$ nel ciclo interno di (6.19), non sappiamo nulla dell'iterata n-sima per $l = 2$ ed allora la $u_2^{(n)}$, effettivamente usata nel ciclo, è pari alla soluzione minima determinata nel passo precedente p_{k-1} in $l = 2$.*

Nel caso $l = N$, l'unico $Q_{k,N} \neq 0$ è $Q_{N,N} = 1$ per cui, considerato che $Q_{N,N-1} = -1$, si ottiene

$$\sum_{k=2}^{N} {g_{\alpha,\lambda}^{(p)}}' \left(\sum_{s=1}^{N} Q_{k,s} u_s^{(n)} \right) Q_{k,N} = {g_{\alpha,\lambda}^{(p)}}' \left(\sum_{s=1}^{N} Q_{N,s} u_s^{(n)} \right) = {g_{\alpha,\lambda}^{(p)}}'(u_N^{(n)} - u_{N-1}^{(n)})$$

da cui

$$u_N^{(n+1)} = u_N^{(n)} - \frac{\omega}{2 + 2\lambda^2} \left(2(u_N^{(n)} - d_N) + {g_{\alpha,\lambda}^{(p)}}'(u_N^{(n)} - u_{N-1}^{(n)}) \right).$$

Osservazione operativa 6.6 *Siccome operativamente nei codici si perdono i risultati intermedi di cicli precedenti, non si ha a disposizione, in questo caso, la* $u_{N-1}^{(n)}$ *che viene sostituita con la* $u_{N-1}^{(p=p_k)}$*, cioè con l'approssimazione della coordinata* $N-1$*-sima del minimo di* $E^{(p)}$ *con* $p = p_k$ *corrente, determinata alla fine del precedente ciclo su* $l = N-1$*.*

Ripetendo le stesse considerazioni si ottiene l'iterata per $2 \leq l \leq N-1$

$$u_l^{(n+1)} = u_l^{(n)} - \frac{\omega}{2+4\lambda^2}\left(2(u_l^{(n)} - d_l) + {g_{\alpha,\lambda}^{(p)}}'(u_l^{(n)} - u_{l-1}^{(n)}) + {g_{\alpha,\lambda}^{(p)}}'(u_l^{(n)} - u_{l+1}^{(n)})\right).$$

Osservazione operativa 6.7 *In questa iterata valgono entrambe le osservazioni precedenti. Quindi la* $u_{l-1}^{(n)}$ *viene sostituita operativamente con l'approssimazione del minimo* $u_{l-1}^{(p=p_k)}$ *ottenuta nel ciclo precedente su* $l-1$*, mentre* $u_{l+1}^{(n)}$ *con l'approssimazione di* $u_{l+1}^{(p=p_{k-1})}$ *ottenuta alla fine del ciclo per* $p = p_{k-1}$ *e* $l+1$*.*

Nel successivo capitolo per il modello di weak membrane viene descritta una rappresentazione funzionante con due array.

Capitolo 7

Algoritmo GNC per il modello weak membrane

L'obiettivo di questo capitolo è quello di descrivere il codice "GNC.c" allegato in fondo, da un punto di vista teorico, partendo dalla implementazione delle istruzioni del tipo (6.18) che regolano l'algoritmo GNC sul modello di weak membrane. Successivamente vengono discussi i risultati dell'applicazione dell'eseguibile GNC.exe su un'immagine 512×512 pixel.

7.1 Implementazione algoritmo GNC

Supponiamo di avere un dominio Ω discretizzato mediante una matrice $n_2 \times n_1$, $(x_{i,j})_{\{1 \leq i \leq n_2, 1 \leq j \leq n_1\}}$. Questa discretizzazione si può trasformare in una che dipende da un solo indice: basta indicare gli elementi della matrice $x_{i,j}$ con l'elemento x_l corrispondente nella matrice (A.2) che ha indici crescenti sulle colonne. Consideriamo poi il modello di weak membrane di $E^{(p)}$ nella forma dedotta da quella generale (6.14) che

diventa considerando la rappresentazione (A.8),

$$E^{(p)}(u) = \sum_{i=1}^{n_1 n_2}(u_i - d_i)^2 + \sum_{k=n_1+1}^{n_1 n_2} g_{\alpha,\lambda}^{(p)}\left(\sum_{s=1}^{n_1 n_2} P_{ks}u_s\right) + \sum_{k=2}^{n_1 n_2} g_{\alpha,\lambda}^{(p)}\left(\sum_{s=1}^{n_1 n_2} R_{ks}u_s\right)$$

Allora l'iterata (6.10) diventa nel caso del modello della weak membrane, utilizzando il risultato (6.15) $\forall l = 1, \ldots, n_1 n_2$

$$\begin{aligned} u_l^{(n+1)} \;=\; u_l^{(n)} - \frac{\omega}{T_l} \;\Bigg(& 2(u_l^{(n)} - d_l) + \sum_{k=n_1+1}^{n_1 n_2} g_{\alpha,\lambda}^{(p)\,\prime}\left(\sum_{s=1}^{n_1 n_2} P_{k,s}u_s^{(n)}\right) P_{k,l} \\ & + \sum_{k=2}^{n_1 n_2} g_{\alpha,\lambda}^{(p)\,\prime}\left(\sum_{s=1}^{n_1 n_2} R_{k,s}u_s^{(n)}\right) R_{k,l}\Bigg) \end{aligned} \tag{7.1}$$

dove ω è definita in (6.13) e T_l ha la forma

$$T_l = 2 + 2\lambda^2\left(\sum_{k=n_1+1}^{n_1 n_2} P_{k,l}^2 + \sum_{k=2}^{n_1 n_2} R_{k,l}^2\right).$$

Ora come per la weak string anche in questo caso l'iterata generale (7.1) ha diverse forme a seconda del punto x_l considerato. Ritornando alla discretizzazione con due indici (i, j) otteniamo 9 zone della matrice in cui l'espressione dell'iterata è diversa. I seguenti risultati vengono ottenuti analizzando le espressioni delle matrici P e Q date nell' appendice A.2.1.

- Posizione $(1, 1)$: $\quad T_{1,1} = 2 + 4\lambda^2$;

$$\begin{aligned} u_{1,1}^{(n+1)} \;=\; u_{1,1}^{(n)} - \frac{\omega}{2+4\lambda^2} \;\Big\{ & 2(u_{1,1}^{(n)} - d_{1,1}) + g_{\alpha,\lambda}^{(p)\,\prime}(u_{1,1}^{(n)} - u_{2,1}^{(n)}) \\ & + g_{\alpha,\lambda}^{(p)\,\prime}(u_{1,1}^{(n)} - u_{1,2}^{(n)})\Big\}. \end{aligned}$$

- Posizione $(1, 2 \le j \le n_1 - 1)$: $\quad T_{1,j} = 2 + 6\lambda^2$;

$$\begin{aligned} u_{1,j}^{(n+1)} \;=\; u_{1,j}^{(n)} - \; & \frac{\omega}{2+6\lambda^2}\Big\{2(u_{1,j}^{(n)} - d_{1,j}) + g_{\alpha,\lambda}^{(p)\,\prime}(u_{1,j}^{(n)} - u_{2,j}^{(n)}) \\ & + g_{\alpha,\lambda}^{(p)\,\prime}(u_{1,j}^{(n)} - u_{1,j-1}^{(n)}) + g_{\alpha,\lambda}^{(p)\,\prime}(u_{1,j}^{(n)} - u_{1,j+1}^{(n)})\Big\}. \end{aligned}$$

- Posizione $(1, n_1)$: $\quad T_{1,n_1} = 2 + 4\lambda^2$;

$$u_{1,n_1}^{(n+1)} = u_{1,n_1}^{(n)} - \frac{\omega}{2+4\lambda^2} \left\{ 2(u_{1,n_1}^{(n)} - d_{1,n_1}) + g_{\alpha,\lambda}^{(p)\prime}(u_{1,n_1}^{(n)} - u_{2,n_1}^{(n)}) + g_{\alpha,\lambda}^{(p)\prime}(u_{1,1}^{(n)} - u_{1,n_1-1}^{(n)}) \right\}.$$

- Posizione $(2 \le i \le n_2 - 1, 1)$: $T_{i,1} = 2 + 6\lambda^2$;

$$u_{i,1}^{(n+1)} = u_{i,1}^{(n)} - \frac{\omega}{2+6\lambda^2} \left\{ 2(u_{i,1}^{(n)} - d_{i,1}) + g_{\alpha,\lambda}^{(p)\prime}(u_{i,1}^{(n)} - u_{i-1,1}^{(n)}) + g_{\alpha,\lambda}^{(p)\prime}(u_{i,1}^{(n)} - u_{i+1,1}^{(n)}) + g_{\alpha,\lambda}^{(p)\prime}(u_{i,1}^{(n)} - u_{i,2}^{(n)}) \right\}.$$

- Posizione $(2 \le i \le n_2 - 1, 2 \le j \le n_1 - 1) \quad T_{i,j} = 2 + 8\lambda^2$;

$$u_{i,j}^{(n+1)} = u_{i,j}^{(n)} - \frac{\omega}{2+8\lambda^2} \left\{ 2(u_{i,j}^{(n)} - d_{i,j}) + g_{\alpha,\lambda}^{(p)\prime}(u_{i,j}^{(n)} - u_{i-1,j}^{(n)}) + g_{\alpha,\lambda}^{(p)\prime}(u_{i,j}^{(n)} - u_{i+1,j}^{(n)}) + g_{\alpha,\lambda}^{(p)\prime}(u_{i,j}^{(n)} - u_{i,j-1}^{(n)}) + g_{\alpha,\lambda}^{(p)\prime}(u_{i,j}^{(n)} - u_{i,j+1}^{(n)}) \right\}. \tag{7.2}$$

- Posizione $(2 \le i \le n_2 - 1, n_1)$: $T_{i,n_1} = 2 + 6\lambda^2$;

$$u_{i,n_1}^{(n+1)} = u_{i,n_1}^{(n)} - \frac{\omega}{2+6\lambda^2} \left\{ 2(u_{i,n_1}^{(n)} - d_{i,n_1}) + g_{\alpha,\lambda}^{(p)\prime}(u_{i,n_1}^{(n)} - u_{i-1,n_1}^{(n)}) + g_{\alpha,\lambda}^{(p)\prime}(u_{i,n_1}^{(n)} - u_{i+1,n_1}^{(n)}) + g_{\alpha,\lambda}^{(p)\prime}(u_{i,n_1}^{(n)} - u_{i,n_1-1}^{(n)}) \right\}.$$

- Posizione $(n_2, 1)$: $\quad T_{n_2,1} = 2 + 4\lambda^2$;

$$u_{n_2,1}^{(n+1)} = u_{n_2,1}^{(n)} - \frac{\omega}{2+4\lambda^2} \left\{ 2(u_{n_2,1}^{(n)} - d_{n_2,1}) + g_{\alpha,\lambda}^{(p)\prime}(u_{n_2,1}^{(n)} - u_{n_2-1,1}^{(n)}) + g_{\alpha,\lambda}^{(p)\prime}(u_{n_2,1}^{(n)} - u_{n_2,2}^{(n)}) \right\}.$$

- Posizione $(n_2, 2 \le j \le n_1 - 1)$: $\quad T_{n_2,j} = 2 + 6\lambda^2$;

$$u_{n_2,j}^{(n+1)} = u_{n_2,j}^{(n)} - \frac{\omega}{2+6\lambda^2} \left\{ 2(u_{n_2,j}^{(n)} - d_{n_2,j}) + g_{\alpha,\lambda}^{(p)\prime}(u_{n_2,j}^{(n)} - u_{n_2-1,j}^{(n)}) + g_{\alpha,\lambda}^{(p)\prime}(u_{n_2,j}^{(n)} - u_{n_2,j-1}^{(n)}) + g_{\alpha,\lambda}^{(p)\prime}(u_{n_2,j}^{(n)} - u_{n_2,j+1}^{(n)}) \right\}.$$

- Posizione (n_2, n_1): $T_{n_2,n_1} = 2 + 4\lambda^2$;

$$u_{n_2,n_1}^{(n+1)} = u_{n_2,n_1}^{(n)} - \frac{\omega}{2+4\lambda^2} \left\{ 2(u_{n_2,n_1}^{(n)} - d_{n_2,n_1}) + g_{\alpha,\lambda}^{(p)\prime}(u_{n_2,n_1}^{(n)} - u_{n_2-1,n_1}^{(n)}) + g_{\alpha,\lambda}^{(p)\prime}(u_{n_2,n_1}^{(n)} - u_{n_2,n_1-1}^{(n)}) \right\}.$$

7.2 Codice per l'algoritmo GNC

Queste iterate vanno inserite all'interno di 4 cicli a cascata del tipo 6.19: $forp = 1,, p_{min}$, for $i = 1, \ldots, n_2$, for $j = 1, \ldots, n_1$, while $|u_{i,j}^{(n+1)} - u_{i,j}^{(n)} \geq prec$, dove sono presenti delle opportune selezioni per individuare ognuna delle 9 regioni su cui rendere attiva la corrispondente iterata. Le iterate viste vengono modificate, nella pratica, all'interno dei cicli, per le stesse motivazioni date nelle osservazioni operative (6.5,6.6,6.7). Un esempio di algoritmo GNC funzionante scritto in linguaggio C è dato nella funzione *reconst()* all'interno del codice GNC.c . In esso si utilizzano due soli array u e *newu* di dimensione $n_1 \times n_2 = 512 \times 512$, dove newu contiene, nella posizione corrente (i, j), gli aggiornamenti della ricostruzione dei dati d al passo $n + 1$, mentre in quelle precedenti contiene le soluzioni ottime (cioè le approssimazioni delle coordinate $u_{i',j'}^{(p)}$ del minimo di $E^{(p=p_k)}$, in quelle successive le soluzioni ottime per $E^{(p_{k-1})}$. Per cui nel codice, ad esempio, l'iterata 7.2 nella zona centrale del dominio $(2 \leq i \leq n_2 - 1, 2 \leq j \leq n_1 - 1)$ è

```
newu[i][j]=u[i][j] - (relax/(2+8*lambda*lambda))*
( 2*(u[i][j]-data[i][j])+dergp(p,u[i][j]-u[i][j+1],cp,q,r)
+ dergp(p,u[i][j]-u[i+1][j],cp,q,r)
+ dergp(p,u[i][j]-newu[i][j-1],cp,q,r)
+ dergp(p,u[i][j]-newu[i-1][j],cp,q,r) );
```

dove *relax* è il parametro di rilassamento ω del tipo (6.13) e

dergp(p,x,cp,q,r) fa le veci di ${g^{(p)}}'_{\alpha,\lambda}(x)$ con q e p i limiti degli intervalli in cui la funzione cambia forma e $cp = c^{(p)} = c^*/p = 1/4p$.

7.3 Descrizione del codice "GNC.c"

Il codice "GNC.c" permette la ricostruzione di un'immagine perturbata da rumore gaussiano uniforme. L'immagine reale pulita utilizzata è salvata in formato asc sul file "lena.txt". Questo programma opera

Figura 7.1: Immagine reale "lena"

sull'immagine perturbata attraverso il modello adattivo discusso nel paragrafo 5.6. All'inizio dell'esecuzione del programma viene data la

possibilità all'utente di sfruttare i risultati ottenuti in una precedente applicazione di ricostruzione continua con modello di weak membrane (vedi lo schema del paragrafo 5.6) oppure di iniziarne una nuova con nuovi parametri α e λ.

Una nuova applicazione consiste nei passi:

1. Apertura del file esistente "lena.txt";
2. Esecuzione della function *addnoisetoin()* che aggiunge il rumore gaussiano alle intensità dell'immagine "lena" (vedi figura 7.2) ottenendo l'array dei dati *datafld*;
3. avvio del primo passo del metodo adattivo, vedi l' osservazione operativa 5.1, quindi della ricostruzione continua dell'immagine che consiste nelle operazioni
   - acquisizione dei parametri α e λ per ottenere un modello di weak continuity che porti ad una ricostruzione quanto più continua e quindi con:
     $$stddeviationdati \simeq h_{media} < h_0 = \sqrt{\frac{2\alpha}{\lambda}}$$
     e immune al rumore
     $$\alpha\lambda > 2\sigma^2$$
   - applicazione dell'algoritmo GNC cioé della function *reconst* sui dati *datafld* con i parametri α e λ acquisiti.
4. memorizzazione dei risultati su file

Una volta che si dispone della ricostruzione continua (Figura 7.3) e quindi alla fine del primo passo del modello adattivo, viene avviato il secondo passo ossia, rispettando lo schema del paragrafo 5.6 e l'osservazione operativa 5.1:

Figura 7.2: Immagine iniziale sporcata da rumore

1. Si acquisisce il numero di zone in cui dividere l'immagine;

2. Si acquisiscono la sensitività η (*eta*) e il parametro di regolarizzazione λ (*lambda*) controllando che verifichino la condizione di immunità alla comparsa di discontinuità dovute alla digitalizzazione
$$\eta \geq \sqrt{\frac{2}{\lambda}};$$

3. Si determina l'intensità media per ogni regione (x, y), la relativa $h_0(x, y) = \eta I_{av}(x, y)$ e il parametro
$$\alpha(x, y) = \lambda * h_0^2(x, y)/2$$

Figura 7.3: Ricostruzione continua con modello weak membrane. I puntini sono spariti.

che viene indicato nel codice con *alfazona*;

4. Infine si applica l'algoritmo GNC utilizzando l'appropriato *alfazona[x][y]* zona per zona, al posto di α. Si esegue la funzione *reconst()* con l'indice finale 2 che avvisa proprio che le iterate nella funzione devono usare $alfazona$. I risultati vengono poi memorizzati su file in formato asc.

7.4 Applicazione del codice sull'immagine "lena"

Discutiamo i risultati ottenuti applicando il codice all'immagine "lena". Consideriamo il rumore gaussiano uniforme di deviazione $\sigma = 40$, ottenendo l'immagine con rumore (figura 7.2). Si osserva che compaiono molti puntini neri sull'immagine che sono proprio l'effetto di perturbazione che il rumore ha sull'immagine iniziale (figura 7.1). Applicando la ricostruzione continua con parametri $\alpha = 120000$ e $\lambda = 40$ che garantiscono che $h_{media} < h_0$ (ricostruzione continua) e l'immunità dal rumore, otteniamo l'immagine in figura 7.3. In questa si nota che scompaiono le discontinuità dovute al rumore (i puntini) ma l'immagine appare sfocata perché è stata forzata la continuità e quindi anche alcune delle discontinuità dell'immagine di partenza vengono regolarizzate rendendo l'intensità più uniforme "a macchie". Ora vediamo come si comporta la ricostruzione finale a seconda della scelta dei parametri

- $\lambda = 4$, $\eta = 0.8$
  Si ottiene una debole regolarizzazione (λ basso) che porta ad una migliore messa a fuoco dell'immagine, solo che, a differenza della ricostruzione continua, h_0 è più basso e quindi le discontinuità più rilevanti (i puntini presenti nelle zone di luce del corpo e del cappello) permangono. (Figura 7.4)

- $\lambda = 16$, $\eta = 0.4$
  Siccome η è più basso di quello della precedente immagine, allora compaiono più discontinuità dovute al pulviscolo del rumore. Comunque il parametro λ alto, che cerca di regolarizzare tutta l'immagine, tende a nascondere l'effetto del rumore uniformando l'immagine presentando, infatti, delle macchie rispetto alla precedente (cioé abbiamo minore nitidezza nitida). (Figura 7.5)

Figura 7.4: Ricostruzione con $\lambda = 4$ e $\eta = 0.8$

- $\lambda = 16$, $\eta = 0.8$
  Stessa regolarità alta della precedente accompagnata da una sensitività maggiore che elimina i puntini neri. Rispetto alla prima (7.4) ho un eliminazione dei puntini sulle zone di luce. Compare una leggera minore nitidezza dell'immagine dovuta alla forte regolarizzazione.(Figura 7.6). Si osserva che tale immagine è abbastanza simile a quella ottenuta con la ricostruzione continua (Figura 7.3) dove sia h_0 che λ sono alte.

Le immagini migliori si ottengono allora per η alto (cioè le 7.4 e 7.6) che elimina gran parte delle discontinuità isolate (i puntini neri).

Figura 7.5: Ricostruzione con $\lambda = 16$ e $\eta = 0.4$

Come abbiamo visto per eliminarli completamente c'è bisogno di una forte regolarizzazione (λ alto) che però porta a una minore nitidezza con la comparsa di alcune "macchie".

Figura 7.6: Ricostruzione con $\lambda = 16$ e $\eta = 0.8$

Appendice A

Appendice

A.1 Discretizzazione dell'energia

Un'espressione dell'energia in forma discreta nel caso 1-dimensionale cioé con $u : \mathbb{R} \to \mathbb{R}$, e con la funzione di regolarizzazione in funzione delle derivate, è

$$E(u,d) = E(u) = \sum_i (u_i - d_i)^2 + g(u_i - u_{i-1}) = \sum_{i=1}^{N} (u_i - d_i)^2 + \sum_{i=2}^{N} g(u_i - u_{i-1}) \tag{A.1}$$

dove $u_i - u_{i-1}$ approssima nel discreto la $u'(x_i)$, con le x_i, per $i \in \{1, \ldots, N\}$, che discretizzano il dominio Ω lineare dell'Hamiltoniana. Mentre nel caso bidimensionale, se il dominio Ω viene discretizzato con una matrice di n_1 colonne e n_2 righe formata dai punti

$$\begin{pmatrix} x_1 & \ldots & x_{n_1} \\ x_{n_1+1} & \ldots & x_{2n_1} \\ \ldots & \ldots & \ldots \\ x_{n_1(n_2-1)+1} & \ldots & x_{n_1 n_2} \end{pmatrix} \tag{A.2}$$

l'espressione dell'energia in forma discreta diventa

$$E(u,d) = \sum_i (u_i - d_i)^2 + g(u_{i+1} - u_{i-1}, u_{i+n_1} - u_{i-n_1})$$

con le $u_i = u(x_i)$. Generalizzando nel caso D-dimensionale se ogni dimensione del dominio D viene discretizzata con n_i posizioni, l'espressione dell'energia diventa

$$E(u,d) = \sum_i (u_i - d_i)^2 + g(u_{i+1} - u_{i-1}, u_{i+n_1} - u_{i-n_1}, \ldots, u_{i+n_1 n_2} - u_{i-n_1 n_2}, \ldots) \tag{A.3}$$

A.2 Condizioni per la convessità

Ora descriviamo delle condizioni sufficienti a garantire la convessità del funzionale dell'energia nel caso 1-dimensionale e D-dimensionale.

Teorema A.1 *Se la derivata seconda di $g(x)$ è maggiore di $-\frac{1}{2}$ allora il funzionale $E(u)$ espressa nella (A.1) è convesso.*

Dim. Una condizione necessaria e sufficiente che garantisce la convessità per la $E(u) = E(u_1, \ldots, u_n)$ è che la sua matrice hessiana H sia definita positiva. Le derivate parziali del primo ordine sono:

$$\frac{\partial E}{\partial u_1} = 2(u_1 - d_1) - g'(u_2 - u_1)$$

$$\frac{\partial E}{\partial u_i} = 2(u_i - d_i) + g'(u_i - u_{i-1}) - g'(u_{i+1} - u_i), \text{ per i} = 2, \ldots, \text{N} - 1$$

$$\frac{\partial E}{\partial u_N} = 2(u_N - d_N) + g'(u_N - u_{N-1});$$

da cui otteniamo

$$\frac{\partial^2 E}{\partial u_j \partial u_i} = \begin{cases} 0 & j \neq i, i+1, i-1 \\ -g''(u_{i+1} - u_i) & j = i+1 \\ -g''(u_i - u_{i-1}) & j = i-1 \\ 2 + g''(u_i - u_{i-1}) + g''(u_{i+1} - u_i) & j = i \neq 1, N \\ 2 + g''(u_2 - u_1) & j = i = 1 \\ 2 + g''(u_N - u_{N-1}) & j = i = N \end{cases} \tag{A.4}$$

Allora avendo posto

$$g_i'' = g''(u_i - u_{i-1})$$

la condizione

$$x^t H x > 0 \ \forall x \in \mathbb{R}^N$$

che garantisce che la matrice hessiana H di g è definita positiva risulta essere equivalente alla

$$\sum_{i=1}^{N} 2x_i^2 + \sum_{i=2}^{N} (x_i - x_{i-1})^2 g_i'' > 0.$$

Grazie all'ipotesi avremo $g_i'' > -\frac{1}{2}$ e quindi

$$\begin{aligned} \sum_{i=1}^{N} 2x_i^2 + \sum_{i=2}^{N} (x_i - x_{i-1})^2 g_i'' \quad &> \\ \frac{1}{2}\left(\sum_{i=1}^{N} 4x_i^2 - \sum_{i=2}^{N} (x_i - x_{i-1})^2 \right) \quad &> \ 0 \end{aligned}$$

perché vale sempre

$$\sum_{i=1}^{N} (2x_i)^2 \geq \sum_{i=2}^{N} (x_i - x_{i-1})^2$$

per ogni $x \in \mathbb{R}^N$. □

Nel caso D-dimensionale considerato il funzionale dell'energia nella forma (A.3) vale invece il seguente

Teorema A.2 *Se la matrice hessiana di $E(u)$ ha gli autovalori tutti maggiori di $-\frac{1}{2D}$ allora il funzionale $E(u)$ è convesso. Basta che un solo autovalore è minore di $-\frac{1}{2}$ per avere la non convessità del funzionale.*

Per la dimostrazione si veda [3]. □

Una condizione più generale che garantisce la convessità per il modello weak string (A.1) è data dalla seguente

Teorema A.3 *Sia $g(x)$ definita in $\mathbb{R}$ e Q una matrice reale $N \times N$ così definita $Q = (Q_{k,i})$ con $\forall\ i,\ k = 1, \ldots, N$*

$$Q_{k,i} = \begin{cases} 1 & se\ k = i \\ -1 & se\ k = i+1 \\ 0 & se\ k \neq i, i+1 \end{cases} \tag{A.5}$$

Allora se $\exists c^* > 0$ tale che $\forall x \in \mathbb{R}$

$$g''(x) \geq -c^*$$

e se v_{max}, il più grande autovalore di $Q^T Q$, verifica

$$v_{max} \leq \frac{2}{c^*}$$

si ha che la funzione $E(u_1, \ldots, u_n) = E(u)$ nel modello (A.1) è convessa.

Dim. Iniziamo con il considerare la matrice Hessiana $H = (H_{i,j})$ di $E(u)$ che ha gli elementi espressi nella (A.4); essi possono essere rappresentati anche nella forma equivalente compatta

$$H_{i,j} = 2I_{i,j} + \sum_{k=2}^{N} g''(u_k - u_{k-1}) Q_{k,i} Q_{k,j}. \tag{A.6}$$

Avremo la tesi se H è definita positiva. In realtà basterà provare la definita positività nel caso particolare della matrice H' del tipo (A.6) dove $\forall x\ \overline{g}''(x) = -c^*$; come viene provato di seguito.

Data una $\overline{E} = D + \sum_i \overline{g}(u_i - u_{i-1})$ convessa con $\overline{g}$ che verifica $\overline{g}''(x) = -c^*$, varrà l'equivalenza

$$E(u) \text{ è convessa} \Longleftrightarrow \mathrm{E(u)} - \overline{\mathrm{E}}\mathrm{(u)} \text{ è convessa}$$

perché la somma di funzioni convesse è ancora convessa. Quindi verifichiamo ora se

$$E(u) - \overline{E}(u) = \sum_i g(u_i - u_{i-1}) - \overline{g}(u_i - u_{i-1})$$

è convessa, cioè se la sua matrice hessiana H'' è definita positiva. Ossia bisogna provare che $\forall y \in \mathbb{R}^N, y \neq 0$

$$y^T H'' y = \sum_{i=2}^{N} (y_i - y_{i-1})^2 [g''(u_i - u_{i-1}) - \overline{g}''(u_i - u_{i-1})] > 0$$

Ricordando l'ipotesi $g''(t) \geq c^*$ otteniamo che tale condizione è soddisfatta. Quindi $E(u)$ sarà convessa non appena proviamo che lo è $\overline{E}(u)$ cioé che la sua matrice hessiana H' è definita positiva. La matrice $H' = (H'_{i,j})$ con

$$H'_{i,j} = 2I_{i,j} + \sum_{k=2}^{N} c^* Q_{k,i} Q_{k,j} = 2I - c^* Q^T Q$$

ha gli autovalori del tipo $\mu = 2 - c^* v$ con v un autovalore di $Q^T Q$ che ha tutti gli autovalori reali. Allora, se il minimo autovalore μ_{min} di H' è positivo, avremo che H' è definita positiva. Il minimo autovalore è $\mu_{min} = 2 - c^* v_{max}$ che è positivo per l'ipotesi $v_{max} > \dfrac{2}{c^*}$, per cui $\overline{E}(u)$ è convessa. □

A.2.1 Un'altra espressione del modello weak membrane

Qui di seguito viene determinata un'espressione del modello della weak membrane secondo B&Z (6.2) che, mediante un opportuna rappresen-

tazione discreta dei punti, rientra nella categoria dei funzionali del tipo (6.5), che permettono di sfruttare una condizione sufficiente per la convessità (Proposizione 6.1.1) simile a quella per la weak string.

Supponiamo che il dominio venga discretizzato in $n_1 \cdot n_2$ elementi:

$$(x_{i,j}) \text{ con } \begin{matrix} i = 1, \dots n_2 \\ j = 1, \dots n_1 \end{matrix}$$

Scomponiamo il termine di regolarità in due sommatorie ottenendo

$$E(u) = D + \sum_{i,j} g_1(u_{i,j} - u_{i-1,j}) + \sum_{i,j} g_1(u_{i,j} - u_{i,j-1}),$$

dove g_1 è tale che $\forall x, y \quad g_1(x) + g_1(y) = g(x, y)$. Costruiamo ora il vettore u di $\mathbb{R}^{n_1 n_2}$, sfruttando la matrice del dominio del tipo (A.2)

$$u = (u_{11}, u_{12}, \dots, u_{1n_1}, u_{21}, \dots, u_{2n_1}, \dots\dots, u_{n_2 1}, \dots, u_{n_2 n_1}). \qquad \text{(A.7)}$$

Allora otteniamo

$$E(u) = D + \sum_k g_1(\sum_l P_{kl} u_l) + \sum_k g_1(\sum_l R_{kl} u_l) \qquad \text{(A.8)}$$

dove $P, R \in \mathbb{R}^{n_1 n_2 \times n_1 n_2}$ sono due matrici bidiagonali così definite

$$P = \begin{matrix} 1, \dots\dots\dots, n_1 + 1 \\ \end{matrix}
\begin{pmatrix}
\overbrace{1, 0, \dots\dots, 0}, 0, \dots & & \dots\dots, 0 \\
0, 1, 0, \dots\dots & & \dots\dots, 0 \\
\dots, \ddots, \dots\dots & & \dots\dots \\
-1, 0, \dots, 0, 1, 0, \dots & & \dots\dots, 0 \\
0, \ -1, \dots\dots\dots, 0, 1, 0, \dots & & \dots\dots, 0 \\
\dots\dots, \ddots, \dots\dots, \ddots, \dots & & \dots\dots \\
\dots\dots\dots\dots, \ddots, \dots\dots\dots, \ddots & & \dots\dots \\
 & & \\
 & & \\
 & & \\
0, \dots\dots & & \dots, 0, -1, 0, \dots\dots, 0, 1
\end{pmatrix}$$

$$R = \begin{pmatrix} 1, 0, \ldots\ldots & \ldots\ldots, 0 \\ -1, 1, 0, \ldots & \ldots\ldots, 0 \\ 0, -1, 1, 0, \ldots & \ldots\ldots, 0 \\ \ldots\ldots \ddots, \ddots, \ldots & \ldots\ldots\ldots \\ \ldots\ldots\ldots\ldots, \ddots, \ddots, \ldots & \ldots\ldots\ldots \\ 0, \ldots\ldots\ldots & \ldots, -1, 1 \end{pmatrix}$$

A.3 Schema di passaggio da norme robuste a funzioni di penalità

Per il riferimento e il significato delle funzioni usate si veda il paragrafo 4.3. Sia data la norma robusta $\rho(x)$ allora si determina la $psi(x)$ equivalente seguendo i passi:

1. Si definisce $\phi(\omega) = \rho(\sqrt{\frac{\omega}{\tau}})$ dove τ è un opportuno parametro scalare che semplifica l'espressione di ϕ;

2. si calcolano le derivate $\phi'(\omega)$ e $\phi''(\omega)$ e si verifica se

$$\lim_{\omega \to 0} \phi'(\omega) = 1,$$

$$\lim_{\omega \to \infty} \phi'(\omega) = 0,$$

$$\phi''(\omega) < 0$$

   altrimenti si esce dallo schema;

3. si definisce il line process $z = \phi'(\omega)$

4. si pone $\omega = \phi'^{-1}(z)$, cioé si risolve l'equazione $\phi'(\omega) = z$

5. infine si definisce la funzione di penalità

$$\psi(z) = \phi(\omega) - z\omega = \phi((\phi')^{-1}(z)) - z(\phi')^{-1}(z).$$

A.4 Formula di Eulero Lagrange

Teorema A.4 *Sia $f \in \mathcal{C}^1([a,b] \times I\!R \times I\!R)$, consideriamo un insieme di funzioni $\mathcal{U}$ e un funzionale $J : \mathcal{U} \to I\!R$ cosí definito*

$$J(u) = \int_a^b f(x, u(x), u'(x))dx,$$

allora le $u(x)$ minime o massime per J soddisfano la cosiddetta equazione di Eulero-Lagrange:

$$\frac{d}{dx} f_{u'}(x, u(x), u'(x)) - f_u(x, u(x), u'(x)) = 0.$$

La formula dell'energia senza la continuità debole (cioé del tipo $D + S$) della stringa è

$$E(u, d) = \int_a^b ((u(x) - d(x))^2 \, dx + \lambda^2 \int_a^b u'(x)^2 \, dx, \tag{A.9}$$

essa forza la continuità nella ricostruzione e quindi la u che la minimizza sarà una ricostruzione ovunque continua della stringa reale. Determiniamo questa ricostruzione continua, iniziando col determinare la relativa equazione di Eulero Lagrange.

Proposizione A.4.1 *L'espressione della equazione di Eulero Lagrange associata al funzionale dell'energia (A.9) è*

$$u(x) - \lambda^2 u''(x) = d(x). \tag{A.10}$$

Dim. La $f(x, u, u')$ ha le derivate parziali

$$f_u(x, u, u') = 2(u - d),$$

$$f_{u'}(x, u, u') = 2\lambda^2 u'$$

per cui l'equazione di Eulero Lagrange è

$$\frac{d}{dx}(2\lambda^2 u'(x)) - 2(u(x) - d(x)) = 2\lambda^2 u''(x) - 2(u(x) - d(x)) = 0$$

da cui otteniamo

$$\lambda^2 u''(x) - (u(x) - d(x)) = 0$$

e quindi la tesi. □

Ora, per il teorema A.3 la $\overline{u}(x)$ che minimizza l'energia (A.9) fra tutte quelle che soddisfano la condizione $u'|_{a,b} = 0$, che individua uno spazio $\mathcal{U}$, verificherà l'equazione (A.10) e ció porta all'espressione della ricostruzione continua

$$\overline{u}(x) = \int_a^b G(x, x')d(x')dx', \tag{A.11}$$

dove $G(x, x')$ è la funzione di Green (si veda [2]) che nel caso di $a \to -\infty$ e $b \to +\infty$ è

$$G(x, x') = \frac{1}{2\lambda} \exp\left(-\frac{|x - x'|}{\lambda}\right).$$

Si può provare inoltre che il valore minimo dell'energia (A.9) cioè quella valutata in $\overline{u}(x)$ del tipo (A.11), ha la forma

$$E = \int_a^b d(x)(\overline{u}(x) - d(x))dx.$$

□

A.5 Variabili aleatorie gaussiane

Una variabile aleatoria X si dice gaussiana se ha la distribuzione normale di media μ e varianza σ^2 ($X \sim N(\mu, \sigma^2)$), ossia se la sua funzione di densità di probabilità è

$$f(x) = P(X = x) = \frac{1}{\sigma\sqrt{2}} \exp\left(-\frac{1}{2}\left(\frac{x - \mu}{\sigma}\right)^2\right).$$

Proposizione A.4.2 *Se* $X \sim N(\mu, \sigma^2)$ *allora* $P(|X - \mu| < 2\sigma) = 0.9546$.

Quindi una variabile aleatoria gaussiana R di media 0 e deviazione standard $\frac{\sigma}{\sqrt{2}}$ è del tipo $N(0, \frac{\sigma^2}{2})$ e quindi per la proposizione precedente nel 95% dei casi si ha che $|R| < 2\frac{\sigma}{\sqrt{2}} = \sqrt{2}\sigma$ e quindi

$$2\sigma^2 > R^2.$$

Allora dato un $\alpha > 2\sigma^2$ avremo che nel 95% dei casi

$$\alpha > R^2.$$

Bibliografia

[1] Young, Gerbrands, Van Vliet *Foundamentals of Image Processing*, Netherlands, Delft University of Technology, 1998.

[2] A. Blake, A. Zisserman, *Visual Reconstruction* MIT Press, 1987.

[3] Mads Nielsen, *Surface Reconstruction: GNCs and MFAs*, Rapport de recerche N. 2354, September 1994, Unité de recerche INRIA Sophia-Antipolis Cedex(FRANCE).

[4] M. J. Black, A. Rangarayan, *On the unification of Line Processes, Outlier Rejection, and Robust Statistics with Applications in Early Vision*, International Journal of Computer Vision 19(1), 57-91 (1996), 1996 Netherlands.

[5] Prof. David J. Heeger *Notes on Regularization and Robust Estimation* Psych 267/CS 348D/EE 365, September 15, 1998.

Codice “GNC.c”